

TARTU ÜLIKOOL  
Keemia Instituut  
Keemia õppekava

Karl Normak

# GDNF mimeetikute $GFR_{\alpha 1}$ -RET aktiveerimise eksperimentaalne uurimine ja QSAR modelleerimine

Bakalaureusetöö (12 EAP)

Juhendajad: Kaido Tamm, *Ph. D*  
Lauri Sikk, *Ph. D*

Tartu 2017

## Experimental study and QSAR modelling of GFR $\alpha$ 1-RET activation by GDNF mimetics

Karl Normak

### Abstract:

This thesis provides experimental and theoretical research of the GDNF family ligands, their mechanism of action, activated pathways and the effects on the organism. Glial cell-line derived neurotrophic factors (GDNFs) are small proteins, which conduct their signalling through GFR $\alpha$  receptors.[1] In the case of difficult neurodegenerative diseases i.e. Parkinson's and ALS, effective and long-term treatment requires the revival of lost neurons. GDNF is capable of supporting the survival of dopaminergic neurons, which degenerate in Parkinson's disease and spinal motoneurons, which are affected by ALS. [2, 3]

70 compounds were tested on a high-throughput assay to assess their capability to activate GFR $\alpha$ -RET complex. The structures of the compounds were analysed to determine the structure motifs responsible for the activation of the complex. Obtained experimental data was used to develop a QSAR model for further prediction purposes.

Both internal and external validation confirmed, that the developed four parameter QSAR model is capable of predicting and describes the given property. The regression correlation coefficient ( $R^2$ ) is 0,77 and cross-validated correlation coefficient ( $R_{cv}^2$ ) is 0,73. Taking into consideration the complex nature of the observed property, the developed model has good statistical stability.

The descriptors in the developed model are connected to the interactions between the compound and the target GFR $\alpha$ -RET complex. The descriptors in the model describe the shape, flexibility, charge distribution and electron structure of the molecule.

The obtained model reliably describes the ability to activate the GFR $\alpha$ -RET complex and is statistically stable to predict the activities of unmeasured compounds.

### Keywords:

QSAR, GDNF

**CERCS:** P410 Theoretical chemistry, quantum chemistry

# GDNF mimeetikute GFR $\alpha$ 1-RET aktiveerimise eksperimentaalne uurimine ja QSAR modelleerimine

Karl Normak

## Lühikokkuvõte:

Käesolevas töös vaadeldakse ülevaatlikult GDNFi perekonna ligandide toimemehhanisme, nende aktiveeritud signaaliradu ja mõju organismi toimimisele. Gliia rakuliinidest saadud neurotroofiline faktor (GDNF) on väike proteiin, mis kannab oma signaali läbi GFR $\alpha$  retseptorite.[1] Raskete neuroloogiliste haiguste nagu Parkinsoni tõve või ALS efektiivseks ja püsivaks raviks on vajalik neuronite taaselustamine. GDNF aitab päästa dopaminergilisi närvirakke, mis manduvad Parkinsoni tõvega, ning seljamotoneuroneid, mida mõjutab ALS. [2, 3]

Töö eksperimentaalses osas mõõdeti andmed 70 ühendi võimekuse kohta aktiveerida GFR $\alpha$ -RET kompleksi ning seeläbi pidurdada raku hääbumist. Ühendite struktuure analüüsiti, et selgitada välja millised struktuurimotiivid põhjustavad kompleksi aktivatsiooni. Saadud andmetest arendati GFR $\alpha$ -RET kompleksi aktivatsiooni kirjeldav QSAR mudel.

Mudeli sisemine ja välimine valideerimine kinnitasid, et koostatud neljadeskriptoriline mudel on statistiliselt stabiilne, ennustusvõimeline ja kirjeldab vaadeldavat omadust. Mudeli regressiooni korrelatsioonikoeffitsient ( $R^2$ ) on 0,77, ristvalideeritud korrelatsioonikoeffitsient ( $R_{cv}^2$ ) on 0,73. Mudeli statistilised näitajad on head arvestades asjaolu, et ennustati väga keerulist bioloogilist aktiivsust.

Deskriptorid, mis esinevad parimas neljadeskriptorilises mudelis, on seostatavad ligandi (madalmolekulaarse ühendi) ja sihtmärgi (GFR $\alpha$ -RET kompleksi) vaheliste interaktsioonidega. Mudelis kasutatavad molekulaardeskriptorid kirjeldavad molekulide kuju, painduvust, laengujao- tust ja elektronstruktuuri.

Saadud mudel kirjeldab madalamolekulaarse ühendi GFR $\alpha$ -RET kompleksi aktivatsiooni usaldusväärselt ning on ennustusvõimeline treeningrühmaga sarnastele ühenditele.

## Võtmesõnad:

QSAR, GDNF

CERCS: P410 Teoreetiline ja kvantkeemia

# Sisukord

<b>Lühendid</b>	<b>2</b>
<b>Sissejuhatus</b>	<b>3</b>
<b>1 GDNFi perekonna ligandid</b>	<b>4</b>
1.1 Kliiniline tähtsus . . . . .	4
1.2 GFLide mimeetikud . . . . .	4
1.3 GDNFi perekonna signaali kandumine . . . . .	5
<b>2 Eksperimentaalne osa</b>	<b>7</b>
2.1 GFR $\alpha$ 1-RET agonistidi lutsiferaas kiirsõelumis süsteem . . . . .	7
2.2 Eksperimentaalsed tulemused . . . . .	8
<b>3 Arvutuslik osa</b>	<b>11</b>
3.1 Kvantitatiivse struktuur-aktiivsus sõltuvuse (QSAR) meetod . . . . .	11
3.1.1 Molekulaarsed deskriptorid . . . . .	12
3.2 GFR $\alpha$ 1-RET kompleksi aktiveerimise QSAR mudel . . . . .	13
3.2.1 Eksperimentaalsed lähteandmed . . . . .	13
3.2.2 QSAR tarkvara ja struktuuride optimeerimine . . . . .	13
3.2.3 QSAR algoritm ja mudeli arendamine . . . . .	13
3.3 Tulemused ja analüüs . . . . .	16
3.3.1 Mudeli interpretatsioon: Deskriptorite analüüs . . . . .	17
3.3.2 Mudeli valideerimine . . . . .	19
3.3.2.1 GFR $\alpha$ 1-RET aktivatsiooni QSAR mudeli valideerimine . . .	20
3.3.3 Mudeli kasutusala . . . . .	23
<b>Kokkuvõte</b>	<b>25</b>
<b>Kasutatud kirjandus</b>	<b>26</b>
<b>Lisad</b>	<b>29</b>

## Lühendid

**ALS** amütroofiline lateraalskleroos.

**AM1** Austin Model 1.

**BMLR** parim multilineaarne regressioon (*i.k. best multi-linear regression*).

**BMP7** Luu morfogeneetiline proteiin 7 (*i.k. Bone morphogenetic protein 7*).

**DMEM** Dulbecco modifitseeritud Eagle'i seerum.

**DMSO** dimetüülsulfoksiid.

**GNDF** Gliia rakkuliinidest saadud neurotroofiline faktor (*i.k. Glial cell line derived neurotrophic factor*).

**GFL** GDNF perekonna ligand (*i.k. GDNF family ligand*).

**GFR $\alpha$**  GDNFi perekonna retseptor alfa (*i.k. GDNF family receptor alpha*).

**GPI** glükosüül-fosfoinositool.

**HEPES** 4-(2-hüdroksüetüül)-1piperasiinetaansulfoon hape.

**HOMO** kõrgeim hõivatud molekulaarorbitaal (*i.k. highest occupied molecular orbital*).

**LMO** jäta mitu välja (*i.k. Leave many out*).

**LOO** jäta üks välja (*i.k. Leave one out*).

**LUMO** madalaim hõivamata molekulaarorbitaal (*Lowest Unoccupied Molecular Orbital*).

**MAPK** Mitogeeni aktiveeritud proteiin kinaas.

**MOPAC** Molecular Orbital PACKage.

**PI3-kinaas** Fosfatidüülinostiool-4,5-bisfosfaat 3-kinaas.

**PLC- $\gamma$**  Fosfolipaas gamma.

**QSAR** kvantitatiivne struktuur-aktiivsus sõltuvus (*i.k. quantitative structure-activity relationship*).

**RET** transfektsiooni käigus ümber paiknenud (*i.k. REarranged during Transfection*) türosiin kinaasi retseptor.

**TGF- $\beta$**  Transformeeruv kasvufaktor beeta.

## Sissejuhatus

Neurodegeneratiivsete haiguste nagu Parkinsoni tõve või amüotroofiline lateraalskleroosi (ALS) ravimine keskendub peamiselt sümptomite peatamisele, kuid ei tegele patsiendi järjest halveneva neuroloogilise võimekuse taastamisega. Viimasel 30 aastal on peamiseks Parkinsoni tõve ravimiks olnud Levodopa, mis ajus, muudetuna dopamiiniks, peatab ajutiselt Parkinsoni motoorsed sümptomid.[4] Lisaks negatiivsete sümptomite peatamisele peaks närvirakkude efektiivne ravi olema võimeline ka taastama kaotatud neuroloogilised võimed. Selleks on vaja päästa neuronid apoptoosist ja soodustada närvijätkete taastamist.

Gliia rakkuliinidest saadud neurotroofiline faktor (GDNF) on näidanud mitmetes katsetes võimekust päästa närvirakke, mistõttu peetakse teda sobivaks kandidaadiks neuronite massilise surmaga seotud haiguste raviks.[5] GDNF kannab oma signaali rakku kaasretseptor  $GFR\alpha$  ja RET kompleksi abil, mis käivitavad rakus olulised neuritogeneesi signaalirajad.

GDNFi polüpeptiid ei ole sobilike farmakoloogiliste omadustega. Parkinsoni ja teiste neurodegeneratiivsete haiguste mehhanismides on palju teadmata, kuid tulenevalt hematoentsefaalsest barjäärast, mida suured valgulised molekulid nagu GDNF ei suuda läbida on efektiivseks ja odavaks raviks vajalik leida neurotroofiliste faktorite madalmolekulaarsed immiteerijaid, ehk mimeetikuid.[6]

Käesolevas töös antakse ülevaade GDNFi perekonna ligandidest, nende retseptoritest, toimemehhanismidest ja mõjust organismile. Eksperimentaalses osas uuriti 70 ühendi võimet aktiveerida  $GFR\alpha$ -RET kompleksi, st. võimet imiteerida GDNFi. Töö teoreetilises osas arendati saadud andmeseeriat kasutades  $GFR\alpha$ -RET kompleksi agoniste kirjeldav QSAR mudel.

# 1 GDNFi perekonna ligandid

Gliia rakkuliinidest saadud neurotroofiline faktori perekonna ligandid (GFL - *GDNF family ligand*) on osa muutuvate kasvufaktorite- $\beta$  (TGF- $\beta$ ) superperekonnast, kus rohked tsüsteiini jäägid paiknevad teineteise suhtes sarnasel moel.[7] GFLide hulka kuuluvad GDNF, neurturin, artemin ja persephin. Vaatamata madalale aminohapete järjestuse sarnasusele, on kirjeldatud TGF- $\beta$  superperekonna liikmed struktuuriliselt sarnaste konformatsioonidega. Sarnaselt teistele TGF- $\beta$  perekonna liikmetele nagu TGF- $\beta$ 2 ja BMP7, on GDNFi dimeeril tasapinnalised "sõrmed", mis koosnevad suurel hulgal  $\beta$ -lehtedest. GDNFi dimeer on painduv. See omadus tuleb  $\alpha$ -heeliksi paiknemisest painduva rühma otsas, mis võimaldab dimeeri eriosadel liikuda, aga siiski säilitada tugevat kontakti teiste osadega.  $\alpha$ -heelikseid kinni hoidev punkt asub valgu keskosas paiknevas tsüsteiini sõlmes. GDNF toimib homodimeerina ning kuulub tsüsteiini sõlmega valkude perekonda. Tsüsteiini sõlm on struktuuri osa, mis koosneb kuuest tsüsteiini jäägist, mis on omavahel kokku seotud kolme disulfiid sidemega.[8]

GFLid osalevad praktiliselt igas neuroni eluetapis, aga nende täpsed toimemehhanismid on jätkuvalt ebaselged. Teatud närvirakud lähevad üle ühelt GFLilt teisele, olgugi, et need mõlemad GFLid suudavad aktiveerida samu signaaliradu. Samas võib üks GFL erinevates raku elu staadiumides omada kategooriliselt erinevaid bioloogilisi toimeid.[8]

## 1.1 Kliiniline tähtsus

GDNFi peetakse atraktiivseks terapeutiliseks vahendiks Parkinsoni tõve ja amüotroofilise lateraalskleroosi (ALS) raviks.[2, 8] GDNF ja teatud teised madalmolekulaarsed mimeetilisid ühendid on suutelised tugevdama dopamiinseid neuroneid, mida mõjutab Parkinsoni tõbi.[5] Mitmed loomsed ja kliinilised katsed on kinnitanud GDNFi positiivset mõju närvirakkude jätkete kasvule. [5, 6] GDNFi on kasutatud näo motoneuronite päästmiseks ja atroofia vältimiseks ning aksotoomia põhjustatud massilise mootorsete neuronite atroofia ja rakusurma vältimiseks.[5] Lisaks on Parkinsoni tõve modelleerivad loomsed katsed näidanud, et GDNFiga saab pidurdada neurotoksiinide indutseeritud dopamiinsete närvirakkude mandumist.[9]

Ka neurturin, persephin ja artemin omavad positiivset mõju erinevate närvihaiguste korral. Neurturin suudab elavdada basaali eesaju kolinergilisi närvirakke ja seljaaju motoorseid neuroneid, mistõttu nähakse neurturinis potentsiaalset Alzheimeri ja ALS ravimit.[10, 11]

## 1.2 GFLide mimeetikud

GFLide polüpeptiidid ei ole sobivad farmakoloogilised agendid, sest nad pole võimelised läbima hematoentsefaalset barjääri ehk veri-aju barjääri. Seetõttu on ilmnenud vajadus arendada väikese molekuliga agonist, mis on suuteline seonduma GFR $\alpha$ ga ja esile kutsuma RETi aktivatsiooni. GFLide mimeetikud ehk ained, mis käituvad sarnaselt GFLidele, annaksid eeldatavalt võimaluse manustada ravimit süstemaatiliselt. Madalmolekulaarne aine kasutamine oleks

farmakokineetiliselt stabiilsem ning võimaldaks ravi odavnemist. Sünteetiliselt loodud väike molekul võib isegi osutada spetsiifilisemaks kui GFL, mida ta imiteerib. GDNF on võimeline signaali üle kandma ka läbi  $GFR\alpha 2$  ja  $GFR\alpha 3$ , aga madalmolekulaarne ühend on eeldatavalt võimeline aktiveerima spetsiifiliselt ainult  $GFR\alpha 1$ -RET kompleksi, aktiveerides kvantitatiivselt erinevaid RETi reguleeritud signaaliradasid.[6, 12]

Selliseid potentsiaalseid aminohapetest mittekoosnevaid mimeetikuid on juba leitud. [5] Sarnaselt GDNFi toimele kiireneb neuronite jätkete kasv nende mimeetikute kontsentratsioonist sõltuvalt. Edasine uurimine on siiski vajalik, kuna leitud mimeetikud on nõrga afiinsusega. Need teadaolevad mimeetikud on suureks abiks suukaudu manustatava GDNFi retseptori agonisti leidmisel.

### 1.3 GDNFi perekonna signaali kandumine

Kõik GDNFi perekonna liikmed kannavad oma signaali läbi türosiin kinaasi retseptori, mis on saanud oma nime teda kodeeriva geeni ümber paiknemise tõttu transfektsiooni käigus (RET – *rearranged during transfection*).[13] RET on rakumembraani läbiv proteiin, mis erinevalt teistest retseptor türosiin kinaasidest koosneb rakuvälises osas neljast kadheriinlikust üksusest ja rakusiseses osas tüüpilisest türosiin kinaasi domeenist.[14] RET on võimeline aktiveerima erinevaid signaaliradu, sh. olulisi neuritogeneesi MAPK ja neuroülekanne tugevdamise PLC- $\gamma$  ja PI3-kinaasi radu.[8, 15, 16] Erinevalt teistest teadaolevatest retseptor türosiin kinaasidest ei toimu signaali ülekandumine RETi ja GFLi otsesel interaktsioonil, vaid kaasretseptor on vajalik. Need kaasretseptorid on tuntud kui GDNFi perekonna retseptor- $\alpha$  ( $GFR\alpha$  – *GDNF family receptor- $\alpha$* ). Ligandi spetsiifilisuse määrab moodustuv  $GFR\alpha$ -RET kompleks.[8]

Signaali ülekandel interakteerub esmalt GDNF vastava  $GFR\alpha 1$  kaasretseptoriga ning seejärel moodustab kompleksi RETiga. Algse mudeli kohaselt on monomeerne või dimeerne  $GFR\alpha 1$  sihtmärgiks GDNFi dimeerile. Nad moodustavad GDNF- $GFR\alpha 1$  kompleksi, mis seostub kahe RET retseptoriga raku välispinnal. Seostumine rakuvälisele domeenile põhjustab homodimeerisatsiooni, RETi aktivatsiooni ja raku siseses osas türosiinide trans-autofosforülatiooni.[17] Pärast seda seostub RETi rakusisene osa erinevate rakusiseste signaalproteiinidega.[12] Eelduste kohaselt käituvad ka teised GDNFi perekonna ligandid sarnaselt GDNFile. Kõigile neljal GFLil on oma peamine kaasretseptor: GDNF seostub  $GFR\alpha 1$ ga, neurturin (NRTN)  $GFR\alpha 2$ ga, artemin (ARTN)  $GFR\alpha 3$ ga ja persephin (PSPN)  $GFR\alpha 4$ ga.[8]

GPI ankurdatud  $GFR\alpha$  on koondunud lipiidparvedesse ja kompleksi moodustumisel RETiga liigub ka RET samasse lipiidparve. Lipiidne parv paneb interakteeruvad molekulid üksteise lähedale mõlemal pool plasmamembraani ning välistab interaktsioonid molekulidega, mis ei asu lipiidses parves. Eksperimentaalsed andmed näitavad, et see parvedesse koondumine on oluline allavoolu signaaliülekandes. RETi ülekandumine lipiidi parvedesse on oluline seal asuvate signaalproteiinide tõttu, sest RET peab efektiivseks signaaliülekandeks seostuma signaalproteiinidega.[18]



Alternatiivne  $\text{GFR}\alpha$ -RET aktivatsiooni mudel väidab, et aktiveeritud retseptori monomeerid ja aktiveerimata dimeerid eksisteerivad tasakaalus aktiveeritud dimeeridega.[12]  $\text{GFR}\alpha$ -RET kompleksi puhul eksisteerib veel teisigi tasakaale, sest  $\text{GFR}\alpha$  saab interakteeruda RETiga ka ligandi puuduses. Vastavalt on väidetud, et eelvalminud  $\text{GFR}\alpha$ -RET ja heterotetrameersed  $(\text{GFR}\alpha)_2$ -(RET)<sub>2</sub> kompleksid eksisteerivad. Neid komplekse eksisteerib organismis piisaval hulgal, et GFLi mimeetikul oleks otstarbekas sihtida neid komplekse. Selline sihtmärgi valik aitab vähendada RETi üleaktiveerimisest tulenevaid võimalikke kõrvalnähtusid.[6]

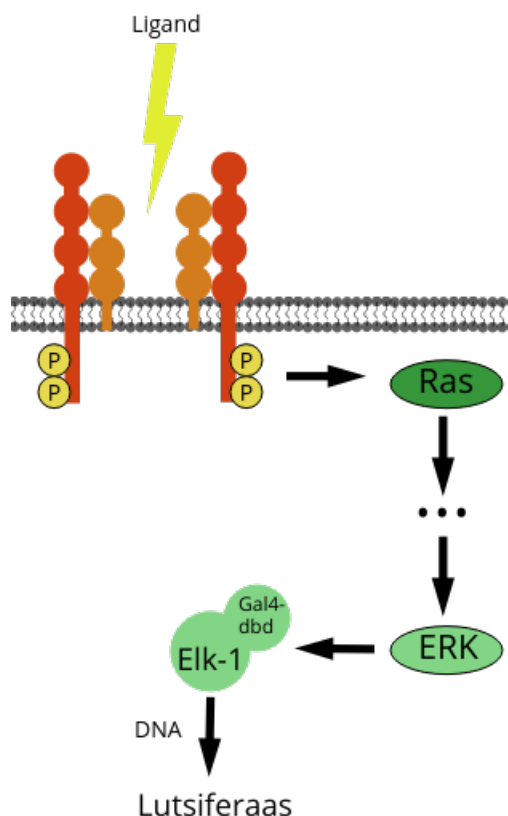
Arvatakse, et türosiini kinaasi retseptorite aktivatsioon toimub väikse molekuli seondumisel teatud kindlatele punktidele juba eelnevalt seondunud retseptori dimeerile. Väikesed molekulid ei ole võimelised ristseostama kahte retseptori molekuli nagu seda teevad bivalentsed ligandid, mistõttu peavad nad põhjustama aktivatsiooni läbi teiste mehhanismide. Üldiselt aksepteeritud arusaam on, et väikese molekuli seostumine kindlatesse punktidesse eelvalminud retseptorkompleksil muudab retseptormolekuli kuju sarnaselt endogeense ligandi seostumisele ning põhjustab retseptori aktivatsiooni. [6]

## 2 Eksperimentaalne osa

### 2.1 GFR $\alpha$ 1-RET agonistidi lutsiferaas kiirsõelumis süsteem

Ühenthite võimet aktiveerida GFR $\alpha$ 1-RET kompleksi uuriti Helsingi Ülikooli Biotehnoloogia Instituudis kiirsõelumissüsteemil prof. Mart Saarma laboris, kus mõõdetakse spetsiaales rakuliinis GFR $\alpha$ 1-RET kompleksi aktivatsiooni kaasnevat bioluminesentsi. Rakuliiniks on MG87 hiire fibroblastid, millesse on transfekteeritud Elk-1/Luc reporter-geeni süsteemi ja GFR $\alpha$ 1-RET kompleksi ekspresseerivad cDNAd. Stabiilset GFR $\alpha$ 1-RET aktivatsiooni lutsiferaasi tootvat rakuliini ning robustset ja tundlikut kiirsõelumissüsteemi on esmakordselt kirjeldanud Sidorova *et. al.*[19].

Metoodika kasutab kõrge tundlikusega reporter-geeni süsteemi, kus GFR $\alpha$ 1-RET kompleksi kontrollitud MAPK signaalirada jõuab lutsiferaasi geeni kontrolliva Elk-1 transkriptaasi faktori posttranslatorse muutuseni. Reporter Elk-1 fosforüleerimine võimaldab tal kinnituda kindlale sünteetilisele promootorile DNA1 ning rakul toota lutsiferaasi. Mida suurem on GFR $\alpha$ 1-RET kompleksi aktivatsioon rakus, seda enam ekspresseeritakse lutsiferaasi tootvat geeni ning seega on suurem ka raku bioluminesents.[19] Reporter-geeni süsteemi kujutav skeem on toodud joonisel 1.



**Joonis 1.** GFR $\alpha$ -RET/Elk-1 vahendatud lutsiferaasi tootmine. Ligandi seostumise järel RET fosforileerub ning käivitab MAPK signaaliraja. ERK fosforileerib Elk-1 aktivatsioonitsentri, mis võimaldab Elk-1 küljes oleval Gal4 DNAle siduval domeenil kinnituda vastavatesse tsentritesse DNA1 ning kutsuda esile lutsiferaasi tootmise.[19]

Stabiilsed GFR $\alpha$ 1, RET ning Elk-1/Luc reporter-geen cDNAGA transfekteritud MG87 hiire fibroblastid plaaditi 96 kaevulistele plaatidele päev enne eksperimenti, igas kaevus 20000 rakku DMEMi, 10% FBSi, 100  $\mu$ g/ml normotsiini, 10  $\mu$ g/ml püromütsiini ja 15 mM HEPESi lahuses (pH 7,2). Järgmisel päeval lisati uuritavaid aineid kontsentratsioonini 5  $\mu$ M DMEMi, 10% FBSi, 100  $\mu$ g/ml normotsiini lahuses. Ühendeid lahjendati DMSOs ning kõikides kaevudes oli DMSO protsent võrdne. Positiivse kontrollina lisati sama kontsentratsiooniga lahuses GDNFi. Seejärel inkubeeriti rakke 24 tundi CO<sub>2</sub> inkubaatoris, et nad saaksid toota lutsiferaasi. Rakud lüüsi kasutades 25  $\mu$ l lüüsi puhvrit (Promega). Valmistati lutsiferaasi testi substraat (Promega), mida lisati 20  $\mu$ l 10  $\mu$ l-ile lüsaadile jää. Luminesentsi loenduriga (Microbeta 2) saadi aktiivsus, mida väljendati signaal-müra ( $S/N$ ) suhtena. Müraks võeti negatiivse kontrolli (proov millele lisati ainult DMSO) luminesents.

Kontsentratsiooni sõltuvuse kindlaks tegemiseks uuriti ühendeid ka kõrgemal, 20  $\mu$ M kontsentratsioonil. Samuti kontrolliti eksperimenti Western blotting meetodil,[20] veendumaks, et kõrgendatud luminesentsi põhjustas GFR $\alpha$ 1-RET kompleksi aktivatsioon (andmeid pole toodud).

## 2.2 Eksperimentaalsed tulemused

Kiirsõelumissüsteemi kasutades mõõdeti 70 ühendi aktiivsus väljendatuna  $S/N$  suhtena. Mõõdetud väärtused jäid vahemikku 0, 83–1, 75. Neli kõige aktiivsemat ühendit on toodud tabelis 1. Kõikide ühendite eksperimentaalselt mõõdetud aktiivsused on toodud tabelis 11. Analüüsimise aktiivsemaid ühendeid lähtudes kesksest tertsiaarsest amiinist. Tabelis 1 toodud ühendite analüüsimisel selgub, et kõigil toodud ühenditel lähtub kesksest amiinist hapnikurikas polaarne haru, teine mittepolaarne haru ja isoksasooli rühma sisaldav kolmas haru. Kõigi aktiivsemate ühendite kolmandas harus, mis sisaldab isoksasooli rühma, on isoksasoolis asendis 3' aromaadne tuum ning asendis 5' on kõigil ühenditel hargnenud või tsükliline tertsiaarne amiin.

Mitteaktiivsetel ühenditel oli, sarnaselt aktiivsetele, suur varieeruvus isoksasooli 5' asendajates, kuid kohati olid isoksasooli rühmad mitteaktiivsel ja aktiivsel ühendil identsed. Tabelis 2 on toodud paaridena üendid, millel olid identsed isoksasooli asendusrühmad, aga väga erinevad aktiivsused. Sellest järeldub, et isoksasooli asendajate täpsest struktuurist ei sõltu oluliselt interaktsiooni tugevus ühendi ja GFR $\alpha$ -RET kompleksi vahel.

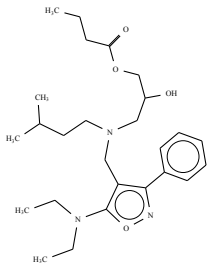
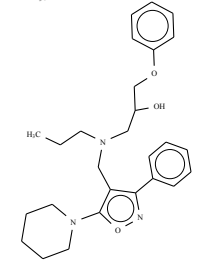
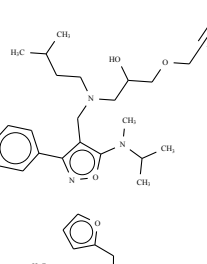
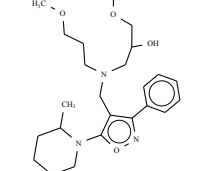
Uurides tabelis 11 ühendeid, siis kõikidel aktiivsematel ühenditel esineb kesksest lämmastikust ühes harus 3 sideme kaugusel hüdroksü ja 4 sideme kaugusel eeter rühm. Nende mõju interaktsioonile ühendi ja GFR $\alpha$ -RET kompleksi vahel on sarnane isoksasooli asendajatele, kus ei leidu selget korrelatsiooni haru ülejäänud struktuuri ja aktiivsuse vahel.

Kõige selgem seos struktuuri ja aktiivsuse vahel ilmneb analüüsides kolmandat mittepolaarset haru. Mõnigaste eranditega on kõik propüüli, isobutüüli või isopentüüli haruga üendid kõige aktiivsemad, nagu on näha tabelis 1 toodud ühenditest. Analüüsides tabelis 11 toodud

ühendeid järgnevad aktiivsuse (furanüül)metüül ja metüület-1,2-diüüleeter. Madala või puuduva aktiivsusega ühenditel esineb sageli tsüklopropüülmetüül rühm või keskse lämmastikuga amiidi moodustav 1,3-bensodioksool-5-karboksamiid rühm.

Kuna viimasena analüüsitud haru pole võimeline andma tugevaid interaktsioone sidumispai-  
ga, järeldame, et üldiselt mittepolaarne haru määrab ühendi sobivuse sidumispai-  
ga ning isoksasooli asendajate üldine struktuur 3 asendis aromaatses tuumaga ja 5 asendis aminorüh-  
maga ning teise haru hüdroksi ja eeter rühmad interakteeruvad GFR $\alpha$ -RET kompleksiga ning  
põhjustavad aktivatsiooni.

**Tabel 1.** Suurima eksperimentaalse aktiivsusega ühendid.

Ühendi struktuur	$S/N$	$M_w$	H doonorid	H akseptorid	$\log P$
	1,75	460	1	7	4,74
	1,74	450	1	6	5,05
	1,71	428	1	6	4,05
	1,69	498	1	7	4,79

Aktiivsemate ühendite vastavuse hindamiseks suukaudsele ravimile analüüsiti neid Lipinski reeglitega.[21] Need reeglid põhinevad statistilistel tähelepanekutel, mis on tehtud tuhandete tuntud ravimite ja ravimisarnaste (*druglike*) ühendite uurimisel. Lipinski esimene reegel ütleb, et ravimikandidaadi molekulmass peaks olema väiksem kui 500. Sellele kriteeriumile vastavad kõik leitud aktiivsemad ühendid. Teine kriteerium, logaritmitud vesi-oktanool jaotuskoefitsient ( $\log P$ ) väiksem kui 5, tagab ravimi imendumise läbi rakumembraani. OpenBabeliga [22, 23]

arvutatud log  $P$  väärtus ei vasta aktiivsusest teisel ühendil Lipinski seatud kriteeriumile, olles marginaalselt suurem kui 5. Reeglile mittevastavust saab ignoreerida, kuna tegemist on madalatasemelise arvutusega, ning erinevus reeglist marginaalne. Ülejäänud ühendid vastavad reeglile. Viimased reeglid seavad piirid vesiniksidemetele, doonoreid vähem kui 5 ning akseptoreid vähem kui 10. Sellega on samuti tagatud aine võime läbida rakumembraani. Samas on oluline, et need numbrid ei oleks liiga väikesed, kuna vesiniksidemed on vajalikud ühendite lahustumiseks. Mõlemale vesiniksideme kriteeriumile vastavad kõik analüüsitud ühendid. Teostatud analüüsi põhjal saab väita, et ühendid on, vastavalt Lipinski reeglitele, ravimisarnased ning sobivad edasiseks ravimi leidmise lähtepunktiks.

**Tabel 2.** Isooksasooli asendajate võrdlus.

Struktuur	$S/N$	Struktuur	$S/N$
	1,71		0,84
	1,75		1,05
	1,75		1,1

## 3 Arvutuslik osa

### 3.1 Kvantitatiivse struktuur-aktiivsus sõltuvuse (QSAR) meetod

Kvantitatiivse struktuur-aktiivsus sõltuvuse (*i.k. Quantitative Structure Activity Relationship* - QSAR) meetodi aluseks on Crum-Browni ja Fraseri püstitatud idee, et aine füsioloogiline aktiivsus on funktsioon tema keemilisest koostisest. Muutes keemilist koostist muutub ka vastavalt bioloogiline aktiivsus.[24] Tänapäevase QSAR analüüsi rajajaks peetakse Corwin Hanschi ja Toshio Fujitat, kes koostasid lihtsa mudeli, mis seostas keemilise ühendi bioloogilise aktiivsuse ning konkreetset ühendit kirjeldava arvvaartuse.[25]

QSAR lähenemist saab üldiselt vaadelda kui andmete analüüsi ja statistiliste meetodite rakendamist bioloogilise või muu aktiivsuse ennustamiseks lähtudes keemiliste ühendite struktuurist.[26]

Klassikalises kvantitatiivses struktuur-aktiivsus sõltuvuses on bioloogiline aktiivsus seostatud mingite füsiko-keemiliste väärtustega, mis tulenevad molekuli struktuurimotiividest. QSAR mudeli koostamiseks kasutatud molekule kirjeldavaid täpse matemaatilise valemiga arvutatud või eksperimentaalselt mõõdetud arvvaartusi nimetatakse deskriptoriteks. Molekuli ehitust kirjeldavad deskriptorid jagunevad rühmadesse: 1. empiirilised (eksperimentaalselt mõõdetud väärtused), 2. kvantkeemilised (näiteks HOMO ja LUMO energiad) ja 3. teoreetilised (näiteks vesiniksideme doonori pindala).[27, 28]

Mudeli koostamise aluseks olevatest deskriptoritest on võimalik saada informatsiooni bioloogiliste, keemiliste ja/või farmakoloogiliste protsesside tagamaadest. Korrektselt koostatud mudel võimaldab ennustada seni mmõõtmata keemiliste ühendite omadusi.[27, 28]

Mitme muutujaga lineaar regressioon analüüsil kooostatud QSAR mudel eeldab, et mingi vastus  $y$  on lineaarkombinatsioon muutujatest  $x_i$ . Mitmeparameetiline mudel on edasiarendus üheparameetrisest lineaarsest sõltuvusest, selliselt, et iga lisatav muutuja vähendaks vea ruutude summat. Mitmeparameetrilise sõltuvuse korral saab sõltuvuse välja kirjutada üldkujul:[28]

$$Y = b + \sum a_i x_i \quad (1)$$

kus  $Y$  on uuritav omadus,  $b$  summaarne vabaliige,  $x_i$  arvutatud molekulaardeskriptor ja  $a_i$  deskriptorile vastav kordaja. Varieerides molekulaardeskriptoreid, nende arvkordejaid ning deskriptorite arvu saame muuta mudeli ennustusvõimet. Rakendades erinevaid algoritmide kriteeriumeid saame optimeerida mudelis esinevaid deskriptoreid.

QSAR mudeli arendamine vajab kolme komponenti: 1. peab olema grupele ühenditele mõõdetud eksperimentaalne väärtus, 2. peavad olema teada selle molekulide grupi struktuurid või omadused (deskriptorid) ning 3. on vaja statistilist meetodit, mis seob omavahel eksperimentaalse ja arvutusliku informatsiooni. Korrektselt koostatud mudelis on optimaalne arv deskriptoreid, mis omavad selgesti tõlgendatavat sisu.[28]

QSAR mudeli esimesest komponendist tuleneb ka mudeli ennustusvõimet piirav faktor.

QSAR mudel ei saa olla statistiliselt täpsem kui esialgne eksperimentaalne info, mida kasutati selle mudeli koostamiseks.

### 3.1.1 Molekulaarsed deskriptorid

Molekuli kajastavaid füsiko-keemilisi deskriptoreid on võimalik leida empiiriliselt või kasutades teoreetilisi ja arvutuslikke meetodeid. Ühendile mõõdetud mingid füsiko-keemilised eksperimentaalsed väärtused või nende põhjal arvutatud väärtused on empiirilised deskriptorid. Molekuli valemist kindlate algoritmidega tuletatud väärtused on aga teoreetilised deskriptorid. Kvantkeemilised deskriptorid erinevad teoreetilisest, kuna põhinevad molekuli lainefunktsioonil.[27]

Kõige lihtsamad teoreetilised deskriptorid on konstitusioonilised, mis väljendavad ühendi keemilist koosseisu. Kindla aatomi koguarv, aromaatsete või muude tsüklite arv või molekulmass on konstitusiooniliste deskriptorite tüüpilised esindajad. Konstitutsioonilised deskriptorid ei vaja arvutamiseks molekuli geomeetrilist struktuuri ega anna ka infot selle kohta.[27, 29]

Topoloogilised deskriptorid kirjeldavad molekuli hargnevust ja aatomitevahelisi sidemeid. Topoloogilised deskriptorid põhinevad molekuli kahedimensionaalsel graafil. Tüüpilisemad topoloogiliste deskriptorite näited on Kier'i ja Hall'i indeksid ning Wiener'i indeks.[27, 29]

Molekuli pindala ja ruumala kirjeldavad geomeetrilised deskriptorid, mille arvutamiseks on vajalik molekuli kolmedimensionaalne struktuur. Geomeetrilised deskriptorid sõltuvad olulisel määral molekuli konformatsioonist, mistõttu on oluline kontrollida konformatsiooni õigsust enne deskriptorite arvutamist. Olulisemad geomeetrilised deskriptorid on molekuli pindala, inertsimoment ning gravitatsiooni indeksid.[27, 29]

Laengute jaotust ning tugevust molekulis kirjeldavad elektrostaatilised deskriptorid, mis põhinevad poolempiirilistel või empiirilistel meetoditel arvutatud osalaengutel. Elektrostaatiliste deskriptorite näideteks on osalaengud, polaarsusparameetrid, polariseeritavus dipoolmoment ja ionisatsioonienergia.[27, 29]

Omaette kategooria moodustavad nii laengute jaotust kui ka molekuli geometriat arvestavad deskriptorid. Laetud osapindala (CPSA-*charged partial surface area*) deskriptoreid kasutatakse molekulidevahelise polaarse vastasmõju kirjeldamiseks. Laetud osapindala deskriptorite näideteks on laengu pindala ning vesiniksideme doonori või akseptori pindala.[27, 29]

Kvantkeemilised deskriptorid arvutatakse lähtudes molekuli kirjeldavast lainefunktsioonist. Kvantkeemilised deskriptorid kirjeldavad molekulisisesid elektroonseid ja tuumade interaksioone. Kvantkeemiliste deskriptorite alaliikideks on molekulaarorbitaalsed ja termodünaamilised deskriptorid.[27, 29]

Molekulaarorbitaalsed deskriptorid kirjeldavad elektronide jaotust molekulaarorbitaalidele. Kõige enam kasutatud molekulaarorbitaalsed deskriptorid on kõrgeima hõivatud molekulaarorbitaali (HOMO) energia, madalaima hõivamata molekulaarorbitaali (LUMO) energia või nende energiatasemete vahe.[27, 29]

Termodünaamilised deskriptorid arvutatakse kasutades kvantekeemilisi meetodeid ja kirjeldavad molekuli termodünaamili omadusi nagu molekuli vibratsiooniline entalpia või entroopia.[27, 29]

### 3.2 GFR $\alpha$ 1-RET kompleksi aktiveerimise QSAR mudel

Lähtudes töö eksperimentaalses osas saadud andmeseeriast arendati QSAR mudel, mis kirjeldab ühendi võimet aktiveerida GFR $\alpha$ 1-RET kompleksi väljendatuna  $S/N$  suhtena ning seeläbi imiteerida endogeenset retseptori ligandi.

#### 3.2.1 Eksperimentaalsed lähteandmed

Töös kasutati QSAR mudeli arendamiseks ja analüüsiks 70 ühendist koosnevat treeningseeriat ja hilisemaks väliseks valideerimiseks 13 ühendist koosnevat andmeseeriat. Uuritavaks omaduseks on GFR $\alpha$ 1-RET kompleksi aktiveerimise võimekus kontsentratsioonil 5  $\mu M$ , mida väljendab  $S/N$  suhe, mille eksperimentaalne väärtus saadi eelnevalt kirjeldatud kiirsõelumissüsteemi kasutades. Suurem  $S/N$  suhe vastab suuremale GFR $\alpha$ 1-RET aktivatsioonile. Eksperimentaalselt saadud aktiivsused, mudeli ennustatud aktiivsused ning nende erinevus on toodud lisades tabelis 11.

Kontsentratsioon 5  $\mu M$  valiti, kuna vähem ühendeid andsid sellel kontsentratsioonil positiivse tulemuse ning eesmärk on leida ühend, mis toimiks võimalikult madal kontsentratsioonil. Ühendid valiti eksperimentaalsetest lähteandmetest nii, et andmeseeria jaotuks vastavalt normaaljaotusele, eksperimentaalsed väärtused oleks kaalutletud ning esineks piisav struktuuriline mitmekesisus.

#### 3.2.2 QSAR tarkvara ja struktuuride optimeerimine

Ühenditel teostati üldine visuaalne kontroll ISISDraw[30] tarkvaraga ja geomeetria eeloptimeerimine HyperChem Professional 7.02 [31] programmiga, kasutades MM+ molekulaarmehaanilist jõuvälja, mis on edasiarendus MM2 jõuväljast.[32] Samuti teostati automaatselt CODESSA PRO[29] tarkvaras ka geomeetria optimeerimine rakendades MOPAC-i [33] poolempiirilisel AM1 (*Austin Model I*) meetodil.[34]

Järgnevalt arvutati konstitutsioonilised, topoloogilised, geomeetrilised, elektrostaatilised, molekulaarorbitaalsed, kvantkeemilised ja termodünaamilised deskriptorid kasutades CODESSA PRO tarkvara. Kokku arvutati igale ühendile 988 deskriptorit.

#### 3.2.3 QSAR algoritm ja mudeli arendamine

Kuna kõigi võimalike deskriptoritega mudelite läbitöötamine võtab ajaliselt väga kaua aega on leiutatud meetodid selle protsessi lihtsustamiseks ja seeläbi kiirendamiseks.[29]



QSAR mudeli leidmiseks kasutati antud töös parima multi-lineaarse regresiooni meetodit (BMLR - *best multi linear regression*). BMLR koosneb järgnevatest sammudest:[29]

1. Ortogonaalsete deskriptorite paaride leidmine.
2. Leitud paaride hulgast luuakse kahedeskriptorilised võrrandid, mille hulgast valitakse kõrgeima korrelatsioonikoefitsiendiga ( $R^2$ ) mudelid. Edasiseks mudeliarendamiseks valitakse korrelatsioonikoefitsiendi järgi 400 parimat võrrandit.
3. Laiendatakse uute deskriptoritega esialgseid 2 deskriptoriga mudeleid seni kuni  $F$ -kriteerium saab väiksemaks kui parimas 2-deskriptorilises korrelatsioonis.
4. Laiendamist korratakse seni kuni deskriptori lisamine ei paranda enam  $F$ -kriteeriumi.  $R^2$  järgi parimad  $N$ -deskriptorilised korrelatsioonid salvestatakse.

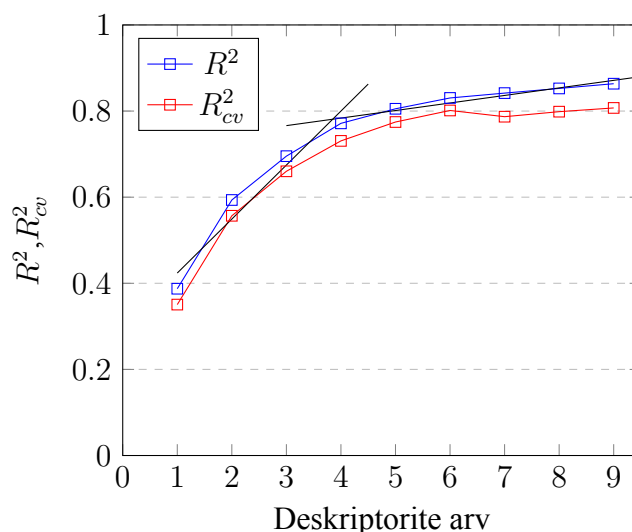
BMLR lahendab edukalt deskriptorite kombinatsioonide üleküllusest tuleneva probleemi valides esimeses sammus piiratud arvu deskriptorite paare. Samas on see ka meetodi puuduseks, kuna esialgne valik on suhteliselt väikese suurusega võrreldes kogu deskriptorite ruumiga.[29]

BMLR meetodit kasutades arendati kahe kuni üheksa deskriptoriga mudelid. Optimaalse deskriptorite arvu määramiseks uuriti korrelatsioonikoefitsiendi ( $R^2$ ) muutumist sõltuvalt deskriptorite arvust mudelis.

Joonisel 2 on toodud deskriptorite arvud mudelis ning vastava deskriptorite arvuga BMLR meetodiga saadud parima mudeli korrelatsioonikoefitsiendid ( $R^2$ ) ning ristvalideeritud korrelatsioonikoefitsiendid ( $R_{cv}^2$ ). Järsku muutust ehk murdepunkti korrelatsioonikoefitsiendi sõltuvuses deskriptorite arvust näeme neljadeskriptorilise mudeli juures. Neljandast deskriptorist alates ei too järgneva deskriptori lisamine kaasa mudeli märkimisväärselt statistilist paranemist. Kui kolme deskriptoriga mudeli  $R^2 = 0,70$  ja nelja deskriptoriga  $R^2 = 0,77$ , ehk  $\Delta R^2 = 0,076$ , siis nelja ja viie deskriptoriga mudeli korral on  $\Delta R^2 = 0,034$ . Võrreldes viienda deskriptori lisamisega kasvab neljanda deskriptori lisamisel  $R^2$  2,3 korda rohkem.

Viienda deskriptori lisamine toob märkimisväärselt väiksema korrelatsioonikoefitsiendi paranemise kui neljanda deskriptori lisamine, mis viitab sellele, et optimaalne deskriptorite arv on saavutatud ja edasine deskriptorite lisamine võib viia mudeli üleparametriseerimiseni. Parimat viiendeskriptorilist mudelit, toodud lisas 10, võrreldes parima neljadeskriptorilise mudeliga, on totaalne soojusmahtuvuse suhe aatomite arvu asendunud sisemise soojusmahtuvuse suhetega aatomite arvu ning lisandunud vesiniksideme doonori pindala (HDSA-2) (MOPAC) deskriptor. Totaalne ja sisemine soojusmahtuvus on oma olemuselt sarnased deskriptorid, nagu on näha ka lisades toodud deskriptorite ja eksperimentaalsete väärtuste graafikutest 8 ja 9, mis korreleeruvad sarnaselt. Vesiniksideme doonori pindala (HDSA-2) (MOPAC) deskriptori koefitsient ja  $t$ -testi väärtus mudelis on oluliselt väiksemad kui teistel deskriptoritel mudelis, mistõttu selle deskriptori kaasamine mudelisse muudab mudeli vähem usaldusväärseks. Lisades toodud

graafikul 11 on näha, et vesiniksideme doonori pindala deskriptor ja eksperimentaalselt saadud väärtused ei korreleeru hästi omavahel, mistõttu selle deskriptori kaasamine mudelis pole põhjendatud.



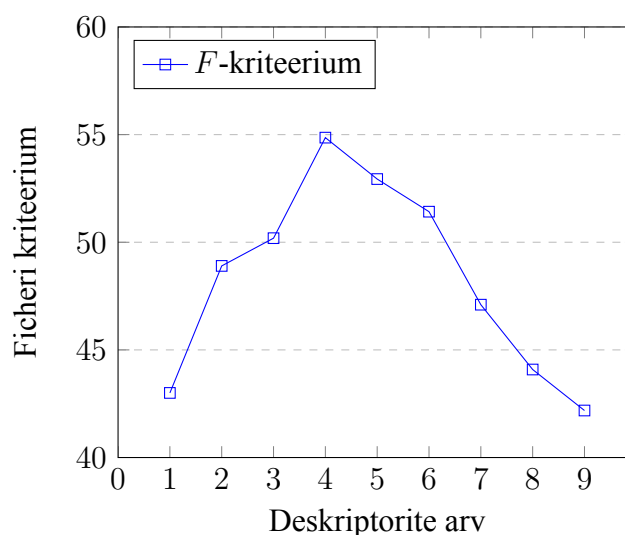
**Joonis 2.** Deskriptorite arv mudelis ja BMLR meetodiga saadud vastavate mudelite korrelatsioonikoeffitsiendid. Sinisega on tähistatud korrelatsioonikoeffitsiendid ( $R^2$ ) ja punasega ristvalideeritud korrelatsioonikoeffitsiendid ( $R^2_{cv}$ ).

Neljadeskriptoriline mudel on statistiliselt oluliselt parema ennustusvõimega kui kolmedeskriptoriline. Sellele viitab neljadeskriptorilise mudeli oluliselt suurem  $R^2$  ja  $R^2_{cv}$  väärtus kui kolmedeskriptorilisel mudelil. Samuti korreleerub deskriptor Zefirovi polaarsus parameeter, toodud lisades joonisel 10, astmeliselt. Parim kolmedeskriptoriline mudel on toodud lisades tabelis 9.

Üheski analüüsitud mudelis ei leidunud olulisi kõrvalekaldujaid. Nende puudumine viitab sellele, et mõõdetud ühendid on struktuuriliselt sarnased.

Neljadeskriptorilise mudeli valikut toetasid ka mudeliseeria Fisheri kriteeriumite väärtused. Fisheri kriteerium võimaldab võrrelda mudeli parameetrite tähtsust, ning teha kindlaks konkreetsed muutujad, mis on mudelis tähendusrikkad ning piisavad. Mudeli parameetrid on piisavad, kui ühegi järgneva deskriptori lisamine ei anna olulist informatsiooni uuritava omaduse kohta.[35] Fisheri kriteerium tavaliselt väheneb parameetri lisamisel siis, kui lisatav parameeter on korrelatsioonis lisaks uuritavale omadusele ka teiste parameetritega.[36] Jooniselt 3 näeme, et neljadeskriptorilise mudeli Fisheri kriteerium on kõige suurem. Seega on neljadeskriptoriline mudel statistiliselt piisav uuritava omaduse kirjeldamiseks.

Ülaltoodut arvesse võttes peeti uuritava omaduse kirjeldamiseks parimaks neljadeskriptorilist mudelit, mida analüüsiti edasi.



**Joonis 3.** Fisheri kriteeriumi sõltuvus deskriptorite arvust mudelis.

### 3.3 Tulemused ja analüüs

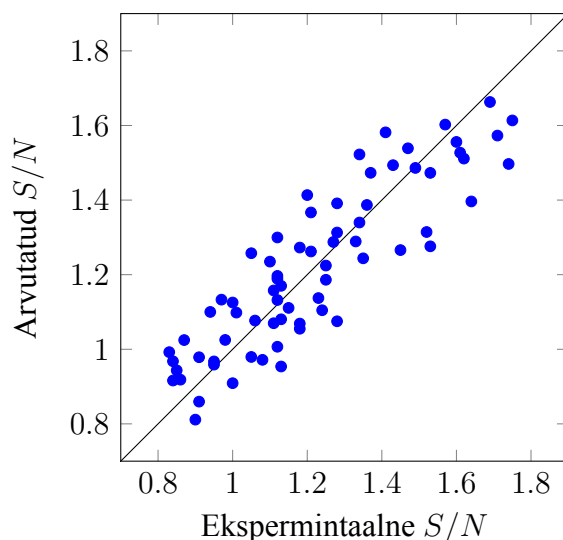
BMLR meetodiga arendatud parim mudel on neljadeskriptoriline, mille deskriptoriteks on termodünaamiline (totaalse soojusmahtuvuse suhe aatomite arvu), molekulaarorbitaalne (HOMO-1 energia), elektrostaatiline (maksimaalne osaline laeng üle kõigi aatomite) ning geomeetiline (ZX varju suhe ZX riskülikusse) deskriptor. Tabelis 3 on toodud mudeli deskriptorid koos vastavate koefitsientide, vabaliikme ja  $t$ -testide väärtustega. Deskriptorid on järjestatud Studenti  $t$ -testi absoluutväärtuse järgi,[36] mis kajastab deskriptorite statistilist olulisust mudelis. Joonisel 4 on toodud parima neljadeskriptorilise mudeliga arvutatud ning eksperimentaalselt mõõdetud  $S/N$  suhte korrelatsioon. Arvestades asjaolu, et ennustatakse keerulist bioloogilist nähtust on mudeli statistilised parameetrid aksepteeritavad ( $R^2 = 0,77$ ;  $s^2 = 0,015$ ) ning kooskõlas mudeli ennustusvõimega ( $R_{cv}^2 = 0,73$ ;  $R_{cvMO}^2 = 0,73$ ).

**Tabel 3.** Neljaparametrilise mudeli deskriptorid ja statistilised näitajad.

Nr.	Deskriptori nimi	Koefitsient	Vabaliige	$t$ -test
1	ZX Shadow / ZX Rectangle	-2,65	0,04	-7,95
2	Max partial charge (Zefirov) for all atom types	6,02	0,31	7,86
3	Tot heat capacity (300K) /natoms	-1,26	0,49	-5,53
4	HOMO-1 energy	0,25	0,31	4,91
Vabaliige			7,09	15,30

$N = 70$ ;  $n = 4$ ;  $R^2 = 0,77$ ;  $R_{cv}^2 = 0,73$ ;  $F = 54,9$ ;  $s^2 = 0,015$

$N$  – ühendite arv,  $n$  - deskriptorite arv,  $R^2$  - korrelatsioonikoefitsient,  $R_{cv}^2$  - ristvalideeritud korrelatsioonikoefitsient,  $F$  - Fisheri kriteerium,  $s^2$  - standardhälbe ruut.



**Joonis 4.** Eksperimentaalsed ja arvutatud  $S/N$  suhte väärtused.

### 3.3.1 Mudeli interpretatsioon: Deskriptorite analüüs

Vastavalt  $t$ -testi väärtustele on kõige olulisem deskriptor mudelis  $ZX$ -peateljestikul molekuli varju pindala suhe samasse teljestikku joonistatud molekuli varju ümbritseva risküliku pindalasse ( $ZX$  Shadow /  $ZX$  Rectangle).  $ZX$  varju suhe  $ZX$  riskülikusse on geomeetriline deskriptor, mis valemina avaldub kujul:[37]

$$S_k^r = \frac{\oint_{(ZX)} (zdx - xdz)}{Z * X} \quad (2)$$

kus integreeritakse üle ühendi kontuuri tasandil, mida defineerivad molekuli  $X$  ja  $Z$  peateljed.[37] Mudelist järeldub, et mida väiksem on see suhe, seda tugevamini aktiveerib ühend  $GFR\alpha$ -RET kompleksi. Optimaalne mimeetik on molekuli  $ZX$  peatelgedes võimalikult kompaktse ehitusega ning hõivab võimalikult väikese ruumi. Deskriptor viitab sellele, et ühend interakteerub  $GFR\alpha$ -RET kompleksiga ruumiliselt piiratud alal, millesse mahtumisel on eelistatud kompaktse ehitusega ühendid. Eksperimentaalseid andmeid analüüsides leiti sarnase olemusega seos ühendite aktiivsuse ja struktuuri vahel, kus kindlas harus kompaktsema rühmaga ühendid olid aktiivsemad. Samuti on märkimisväärne, et tegemist on arendatud QSAR mudeli kõige tähtsama deskriptoriga. See näitab, et deskriptor kirjeldab asjakohast molekuli aktiivsust kirjeldavat omadust.

Marginaalselt tähtsusest  $t$ -testi järgi teine on Zefirovi arvutuseeskirjade järgi arvutatud maksimaalne osalaeng üle kõigi aatomite (*Max partial charge (Zefirov) for all atom types*):[29]

$$Q_{max} = \max(Q^+) \quad (3)$$

Deskriptori väärtus ( $Q_{max}$ ) on võrdne kõige suurema osalaenguga molekulis( $\max(Q^+)$ ).[29]

Deskriptor klassifitseerib ühendid asendajate elektronegatiivsuse ja arvu järgi. Mudeli järgi on eelistatud sellised mimeetikud, kus eksisteerib väga tugeva positiivse laenguga aatom. Deskriptor viitab elektrostaatilistele interaktsioonidele mimeetiku ja  $\text{GFR}\alpha 1\text{-RET}$  kompleksi vahel. Deskriptor on pigem mudelit parandav, kuna deskriptorite väärtuste jaotus on astmeline. Samuti eksisteerivad kaks ühendit, mis häälbivad deskriptori väärtuses olulisel määral.

Totaalne soojusmahtuvus aatomi kohta 300 kelvini juures (*Tot heat capacity (300K) / n-atoms*) on  $t$ -testi järgi mudelis tähtselt kolmas deskriptor. Molekuli soojusmahtuvus sõltub tema rotatsioonilistest, translatsioonilistest ning vibratsioonilistest soojusmahtuvustest, mis omakorda sõltuvad aatomitevahelistest sidemetest ja molekuli ehitusest. Painduvamad, hargnenumad ning ebasümmeetrilised molekulid on suurema soojusmahtuvusega.[33, 38] Mudeli kohaselt on eelistatud ühendid väiksema soojusmahtuvusega aatomi kohta, ehk jäigemad, vähemate vabadusastmetega molekulid. See deskriptor kombineeritult deskriptoriga *ZX Shadow/ZX Rectangle* näitab, et ühend seostub arvatavasti kompaktselt seostumiskohale  $\text{GFR}\alpha 1\text{-RET}$  kompleksil, kus eelistatult seostuvad jäigemad, vähem hargnenud ühendid.

Viimane deskriptor on kõrgeima hõivatud molekulaarorbitaali energia (*HOMO-1 energy*). Deskriptor on arvutatud ühendi kõrgeima hõivatud molekulaarorbitaali lainefunktsioonist ( $\phi_{\text{HOMO}}$ ) rakendades sellele Focki operaatorit ( $\hat{F}$ ):[39]

$$\theta_{\text{HOMO}} = \langle \phi_{\text{HOMO}} | \hat{F} | \phi_{\text{HOMO}} \rangle \quad (4)$$

Kõrge HOMO energiaga ühendid käituvad mudeli kohaselt paremate  $\text{GFR}\alpha 1\text{-RET}$  kompleksi aktivaatoritena. HOMO energia on otseselt seotud molekuli ionisatsiooni potentsiaaliga ning mängib olulist rolli molekuli reaktsioonivõimes ja laenguülekandega komplekside moodustamises. Kvantkeemiliste deskriptoreid ja nende võimet ennustada ainete omadusi on käsitletud M. Karelson *et. al.* artiklis "Quantum-Chemical Descriptors in QSAR/QSPR Studies".[40]

Deskriptorite omavahelise interkorrelatsiooni uuriti vaadeldes deskriptorite omavahelisi korrelatsioonikoefitsiente ( $R^2$ ). Neljadeskriptorilise mudeli deskriptorite omavahelised korrelatsioonikoefitsiendid on toodud tabelis 4.

Deskriptorite omavahelised korrelatsioonikoefitsiendid on tunduvalt väiksemad kui määratud kriteerium ( $R^2 < 0,6$ ). Ühegi deskriptoripaari vahel ei eksisteeri korrelatsiooni. Mudelis olevad deskriptorid on ortogonaalsed ning mudel ise on statistiliselt usaldusväärne.

**Tabel 4.** Neljaparametrilise mudeli deskriptorite omavahelised korrelatsioonikoefitsiendid.

	ZX S/R	Max partial charge	HOMO-1	Tot heat capacity
ZX S/R	1	0,005	0,005	0,013
Max partial charge	0,005	1	0,002	0,253
HOMO-1	0,005	0,002	1	0,242
Tot heat capacity	0,013	0,253	0,242	1

### 3.3.2 Mudeli valideerimine

Enne QSAR meetodiga arendatud mudeli rakendamist on oluline hinnata mudeli statistilist tähtsust ja usaldusväärsust. Selleks, et QSAR mudel oleks ennustusvõimeline peab mudel omama statistilist olulisust ja robustsust, olema valideeritud andmeseeriaga, mida ei kasutatud mudeli koostamiseks, ning omama defineeritud kasutusala. QSAR/QSPR mudeli valideerimine jaguneb meetodikalt kaheks, andmeseeriasisene ja -väline.[28]

Esmaseks mudeli headuse parameetriga on sageli määratlustegur või korrelatsioonikoefitsient  $R^2$ . Mida parem mudel, seda lähemal on  $R^2$  väärtus ühele. Samas tuleb mees pidada, et  $R^2$  on kõigest mudeli sobitavuse parameeter, mistõttu ei kajasta  $R^2$  täielikult mudeli ennustusvõimet. Kõrge  $R^2$  väärtusega QSAR mudel võib olla kehva ennustusvõimega, kui mudelis on vähe vabadusastmeid, muutujate multikolineaarsus jne. [28, 41]

Sisemise valideerimise korral jäetakse mudelist välja üks (LOO - *Leave one out*) või mitu (LMO - *Leave many out*) andmepunkti, millele hiljem ennustatakse väärtused. Ülejäänud andmepunktide deskriptorite ja aktiivsuste vahel arendatakse mudel selleks ennustuseks. Protsessi korratakse jättes välja järgmised andmepunktid. Valideerimisel jälgitakse erinevaid statistilisi kriteeriume, mis peegeldavad mudeli pädevust. Nendeks on korrelatsiooni koefitsiendi ruut ( $R^2$ ), jääkliikmete standardhälve ja ristvalideeritud ruutkorrelatsiooni koefitsient ( $R_{cv}^2$ ). Nende statistiliste parameetrite headus on vajalik ennustusvõimelisele mudelile, aga ei ole piisav mudeli ennustusvõime korrektseks hindamiseks.[28, 42]

Samuti saab lugeda sisemiseks valideerimiseks nn Y-randomiseerimist. Meetod põhineb korduval Y data, ehk mudeli deskriptori väärtuste, juhuslikku järjekorda panemisel, jättes X väärtused, ehk algse eksperimentaalse andmestiku, puutumata. Juhuslik järjekord genereeritakse andes igale ühendile suvalise integraalse väärtuse ühe ja N-i vahel, ning järjestades deskriptori väärtused selle juhusliku integraali järgi. Leitakse uued ennustatud väärtused ning arvutatakse  $R^2$ . Protsessi korratakse mitmeid kordi, sest meetod võib juhuslikkuse tõttu tekitada väga madala või kõrge  $R^2$ -ga tulemi, mis ei pruugi kajastada tegelikku mudeli pädevust. Kui mudel on relevantne, siis selline Y data randomiseerimine põhjustab märkimisväärse languse keskmises  $R^2$  väärtuses mudeli ennustatud ja eksperimentaalsete väärtuste vahel.[43]

Andmeseeriaväline valideerimine hõlmab endast välise andmeseeria kasutamist mudeli üldise täpsuse hindamiseks. Seda välist andmeseeriat nimetatakse sageli testseeriaks ja mudeli koostamiseks kasutatud andmeseeriat treeningseeriaks. Treeningseeria põhjal koostatud mudeli statistilisi parameetreid nagu  $R^2$ , ruutkeskmine viga ja  $R_{cv}^2$  võrreldakse ennustatud testseeria parameetritega. Nende parameetrite erinevus annab üldise hinnangu QSAR mudeli usaldusväärsusele. Lihtsama võrdluse annab välisele andmeseeriale arvutatud ning eksperimentaalse omaduse otsene võrdlemine.[28, 42]

### 3.3.2.1 *GFR $\alpha$ 1-RET* aktivatsiooni QSAR mudeli valideerimine

Juhusliku korrelatsiooni esinemise välistamiseks valideeriti mudel sisemiselt kasutades nn ABC valideerimismeetodit.[44] Ühendid jaotati kolme rühma, nii et igasse rühma kuulus iga kolmas ühend alustades esimesest, teisest või kolmandast ühendist. 1., 4., 7. jne ühend kuulusid rühma A, 2., 5., 8. jne kuulusid rühma B ning 3., 6., 9. jne kuulusid rühma C. Järgnevalt võeti kaks rühma treeningrühmadeks ning kolmas testrühmaks. Treeningrühmaga arendati uus mudel, lastes varieerida ainult parima neljadeskriptorilise mudeli deskriptorite koefitsiente. Näiteks võeti rühma A ja B ühendid treeningrühmaks ning kasutati neid mudeli koostamiseks, mida testiti mudeli koostamisest välja jäetud rühmaga C. Analoogiliselt mängiti läbi kõik treening- ja testrühmade kombinatsioonid. Saadud mudelite korrelatsioonikoefitsiendid on toodud tabelis 5.

**Tabel 5.** ABC valideerimise treening- ja arvutatavad rühmad ning statistilised parameetrid.

Treeningrühm	$R^2$	$s^2$	Arvutatav rühm	$R^2$	$s^2$
$A + B$	0,81	0,013	$C$	0,68	0,019
$A + C$	0,79	0,014	$B$	0,73	0,015
$B + C$	0,72	0,017	$A$	0,84	0,013

$R^2$  - korrelatsioonikoefitsient,  $s^2$  - standardhälbe ruut.

Tabelis 5 toodud andmed viitavad heale korrelatsioonile. Tabelis toodud statistilised näitajad kinnitavad mudeli usaldusväärsust ja ennustusvõimet.

Järgnevalt teostati Y-randomiseerimise test kasutades töö käigus loodud programmi, mis võimaldab kiiresti ja korrektselt korrata Y-randomiseerimise protsessi vajaminev arv kordi. Ühe deskriptori kaupa jaotati juhuslikult ühendite deskriptori väärtused segamini, leiti uued ennustatud väärtused kasutades arendatud mudelit ja kalkuleeriti saadud uute ennustatud ja eksperimentaalsete väärtuste vaheline määratlustegur. Protsessi korrati 1000 korda igale deskriptorile, et välistada juhuslikkusest tulenevat efekti. Selle järel leiti keskmine määratlustegur iga deskriptori randomiseerimisel. Saadud tulemused on toodud tabelis 6.

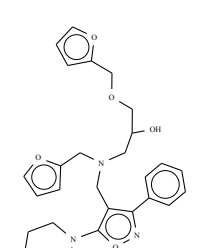
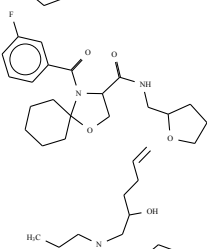
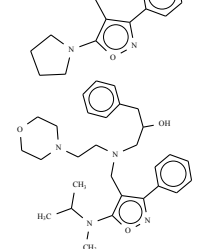
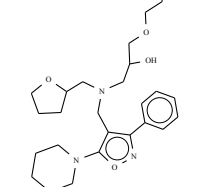

**Tabel 6.** Y-randomiseerimisel saadud keskmised  $R^2$  väärtused

Randomiseeritu deskriptor	keskmine $R^2$
ZX Shadow/ ZX Rectangle	0,31
Max partial charge (Zefirov) for all atom types	0,15
HOMO-1 energy	0,53
Tot heat capacity (300K)/natoms	0,36

Tabelis 6 toodud andmed näitavad, et mudeli deskriptorite randomiseerimisel langeb  $R^2$  märgatavalt. See näitab, et mudel põhineb asjakohastel deskriptoritel, mis ei korreleeru uuritava omadusega juhuslikult. Deskriptori *HOMO-1 energy* randomiseerimisel saadud määratlustegur on kõrgem kui teistel deskriptoritel, kuid siiski märkimisväärselt madalam mudeli määratlustegurist, mistõttu pole alust eemaldada deskriptorit mudelist. Kõrgem  $R^2$  on seletatav deskriptori väiksema panusega mudelisse, mistõttu teised deskriptorid säilitavad mingil määral korrelatsiooni.

Mudeli ennustusvõime kontrolliks valideeriti mudel ka väliselt kasutades ühendeid, mis jäid välja mudeli arendamisest. Vaadeldi arvutatud ja eksperimentaalsete aktiivsuste erinevusi, mis on toodud tabeli 7. Kokku suudeti ennustada piisava täpsusega kõigi 13 ühendi aktiivsused. Testseeria statistilised parameetrid erinevad treeningseeria omadest vähesel määral, testseerias säilis hea korrelatsioonkoefitsient ( $R^2 = 0,80$ ), kuid vähesel määral tõusis dispersioon ( $s^2 = 0,022$ ).

**Tabel 7.** Mudelist välja jäetud ühendite ennustatud aktiivsused

Ühendi struktuur	Eksperimentaalne $S/N$	Arvutatud $S/N$	Erinevus
	1,83	1,68	0,149
	0,87	1,02	-0,151
	2,08	1,71	0,368
	2,08	1,80	0,275
	1,99	1,75	0,235



	1, 95	1,70	0,246
	1, 9	1,65	0,250
	1, 56	1,73	-0,173
	1, 8	1,60	0,203
	1, 78	1,47	0,343
	1, 71	1,37	0,339
	1, 56	1,73	-0,173
	1, 71	1,40	0,307

### 3.3.3 Mudeli kasutusala

Olenemata mudeli statistilistest näitajatest ja valideerimise tulemustest, ei ole reaalne, et mudel on võimeline ennustama kõigi keemiliste ühendite aktiivsusi. Seega on oluline QSAR mudelil defineerida kasutusala, kuhu sobituvate ühendite aktiivsusi on mudel võimeline ennustama usaldusväärselt. [41]

**Tabel 8.** Mudeli kasutusala defineerivad deskriptorite väärtused.

Deskriptor	$D_{min}$	$D_{max}$	Minimaalne	Maksimaalne
ZX Shadow / ZX Rectangle	0,72	0,54	0,78	0,49
Max partial charge (Zefirov) for all atom types	0,09	0,06	0,05	0,10
HOMO-1 energy	-8,70	-10,13	-10,61	-8,23
Tot heat capacity (300K) /natoms	2,04	1,72	2,14	1,62
Ennustatud aktiivsus deskriptorite piirväärtustega:			-0,05	2,27

Vastavalt Eesti standardikeskuse “Kvantitatiivsed Struktuur-Aktiivsus Analüüsid” standardile [45] defineerivad mudeli kasutusala arendamiseks kasutatud treeningseeria ühendid. Kasutusala hõlmab endas ainult neid ühendeid, mis sisaldavad samu aatomeid ja funktsionaalrühmi nagu treeningseeria. Aksepteeritavad deskriptori väärtused ennustatavatele ühenditele on suuremad kui ühe kolmandiku võrra väiksemad treeningseeria minimaalsest deskriptori väärtusest ja väiksemad kui ühe kolmandiku võrra suuremad treeningseeria maksimaalsest deskriptori väärtusest. See nõue on kirja pandud valemina:

$$(D_{min} - \frac{D_{max} - D_{min}}{3}) \leq D \leq (D_{max} + \frac{D_{max} - D_{min}}{3}) \quad (5)$$

kus  $D_{min}$  on treeningseerias vastava deskriptori minimaalne väärtus,  $D_{max}$  – treeningseeria vastava deskriptori maksimaalne väärtus ja  $D$  – ennustatava ühendi deskriptori väärtus. [45]

Mudeli loomiseks kasutatud 4-(aminometüül)-3-fenüül-1,2-oksasool-5-amiini derivaatide deskriptorite maksimaalsed ja minimaalsed väärtused on toodud tabelis 8. Kuna ühendid 18 ja 20 erinesid ülejäänud ühenditest deskriptorite *Max partial charge (Zefirov) for all atom types* ja *Tot heat capacity (300K) /natoms* väärtustes märkimisväärselt, on kasutusala paremaks defineerimiseks mõistlik jätta välja nende ühendite deskriptorite väärtused kasutusala piirväärtuste arvutustest. Samuti on toodud tabelis arvutatud kasutusala maksimaalsed ja minimaalsed väärtused ning nende põhjal mudeliga arvutatud maksimaalne ja minimaalne ennustatav aktiivsus. Kuna deskriptorite *ZX Shadow/ZX rectangle* ja *Tot heat capacity (300K)/natoms* kordajad olid mudelis negatiivsed, siis on pandud minimaalne ja maksimaalne deskriptori väärtus vastavalt.

Tabelis 8 toodud minimaalne mudeli ennustatav piirväärtus on teoreetiline, ning ei oma praktilist tähendust, kuna väga palju ühest väiksem  $S/N$  suhe näitab ühendi toksilisust, mida leitud mudel ei kajasta. Esmases lähenduses saab kasutusala defineerida vaid mudeli deskriptoritega. Täpsema kasutusala määramiseks tuleks arvestada kogu deskriptorite komplekti mis on arvutatud treeningseeria jaoks. Struktuure varieerides ja ennustatud aktiivsuse piirväärtuste ületamisel ei pruugi mudel aktiivsusi usaldusväärselt ennustada.

## Kokkuvõte

Käesolevas töös vaadeldakse ülevaatlikult GFLi perekonna ligandide toimemehhanisme, nende aktiveeritud signaaliradu ja mõju organismi toimimisele. Raskete neuroloogiliste haiguste nagu Parkinsoni tõvi või ALS efektiivseks ja püsivaks raviks on vajalik neuronite taaselustamine, mida on võimalised tegema GFLi perekonna ligandid.

Töö eksperimentaalses osas mõõdeti 70 ühendi võimekust aktiveerida  $GFR\alpha 1$ -RET kompleksi ning seeläbi vältida raku apoptoosi. MG87 hiire fibroblastides, millesse on transfekteeritud Elk-1/Luc reporter-geen süsteem, mõõdeti  $GFR\alpha 1$ -RET kompleksi aktivatsioonil kaasnevat bioluminesentsi. Uuriti ühendite struktuuri ja aktiivsuse sõltuvust ning analüüsiti aktiivsemaid ühendeid Lipinski reeglite järgi nende ravimisarnasuse kindlaks tegemiseks. Ühendite aktiivsus sõltus suuresti ühendite ühe haru ehitusest. Mõõdetud andmetest koostati  $GFR\alpha$ -RET kompleksi aktivatsiooni kirjeldav QSAR mudel.

Sisemine ja välimine valideerimine kinnitasid, et koostatud neljadeskriptoriline mudel on ennustusvõimeline ja kirjeldab vaadeldavat omadust. Mudeli statistilised näitajad on head arvestades asjaolu, et ennustati väga kompleksset bioloogilist nähtust. Mudeli regressiooni korrelatsioonikoefitsient ( $R^2$ ) on 0,77, ristvalideeritud korrelatsioonikoefitsient ( $R_{cv}^2$ ) on 0,73.

Deskriptorid, mis esinevad parimas neljadeskriptorilises mudelis, on seostatavad ligandi (madalmolekulaarse ühendi) ja sihtmärgi ( $GFR\alpha$ -RET kompleksi) vaheliste interaktsioonidega. Mudelis kasutatavad molekulaardeskriptorid kirjeldavad molekulide kuju, painduvust, laengujaotust ja elektronstruktuuri.

Eelnevale tuginedes saab väita, et koostatud neljadeskriptoriline mudel kirjeldab madalamolekulaarse ühendi  $GFR\alpha$ -RET kompleksi aktivatsiooni usaldusväärselt ning mudel on ennustusvõimeline treeningrühmaga sarnastele ühenditele, mis sobivad töös defineeritud mudeli kasutusalas.

Edasine ravimiarendus antud töös kasutatud ühenditele on lõpetatud, kuna leitud ühendid polnud piisavalt aktiivsed.

## Kasutatud kirjandus

- (1) Cacalano, G.; Farin, I.; Wang, L.; Hagler, K.; Forgie, A.; Moore, M.; Armanini, M.; Phillips, H.; Ryan, A. M.; Reichardt, L. F.; Hynes, M.; Davies, A.; Rosenthal, A.; Davies, A.; Rosenthal, A.  $GFR\alpha 1$  Is an Essential Receptor Component for GDNF in the Developing Nervous System and Kidney. *Neuron* **1998**, *21*, 53–62.
- (2) Henderson, C. E.; Phillips, H. S.; Pollock, R. A.; Davies, A. M.; Lemeulle, C.; Armanini, M.; Simpson, L. C.; Moffet, B.; Vandlen, R. A. GDNF: A Potent Survival Factor for Motoneurons Present in Peripheral Nerve and Muscle. *Science* **1994**, *266*, 1062–1064.
- (3) Lin, L. H.; Doherty, D. H.; Lile, J. D.; Bektesh, S.; Collins, F. GDNF: A Glial Cell–Line Derived Neurotrophic Factor for Midbrain Dopaminergic Neurons. *Science* **1993**, *260*, 1130–1132.
- (4) National Collaborating Centre for Chronic Conditions, Royal College of Physicians, *Parkinson's Disease: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care*, London, UK, 2006; Vol. 35.
- (5) Tokugawa, K.; Yamamoto, K.; Nishiguchi, M.; Sekine, T.; Sakai, M.; Ueki, T.; Chaki, S.; Okuyama, S. XIB4035, a novel nonpeptidyl small molecule agonist for  $GFR\alpha-1$ . *Neurochem. Int.* **2003**, *42*, 81–86.
- (6) Bessalov, M. M.; Saarma, M. GDNF family receptor complexes are emerging drug targets. *Trends Pharmacol. Sci.* **2007**, *28*, 68–74.
- (7) Ibáñez, C. F. Emerging themes in structural biology of neurotrophic factors. *Trends Neurosci.* **1998**, *21*, 438–444.
- (8) Airaksinen, M. S.; Saarma, M. the Gdnf Family: Signalling, Biological Functions and Therapeutic Value. *Nat. Rev. Neurosci.* **2002**, *3*, 383–394.
- (9) Grondin, R.; Gash, D. M. Glial cell line–derived neurotrophic factor (GDNF) – a drug candidate for the treatment of Parkinson's disease. *J. Neurol.* **1998**, *245*, 35–42.
- (10) Garcès, A.; Livet, J.; Grillet, N.; Henderson, C. E.; Delapeyrière, O. Responsiveness to Neurturin of Subpopulations of Embryonic Rat Spinal Motoneuron Does Not Correlate With Expression of  $GFR\alpha 1$  or  $GFR\alpha 2$ . *Dev. Dyn.* **2001**, *220*, 189–197.
- (11) Golden, J. P.; Milbrandt, J.; Johnson, E. M. Neurturin and persephin promote the survival of embryonic basal forebrain cholinergic neurons in vitro. *Exp. Neurol.* **2003**, *184*, 447–455.
- (12) Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **2000**, *103*, 211–225.
- (13) Takahashi, M.; Ritz, J.; Cooper, G. M. Activation of a Novel Human Transforming Gene, RET, by DNA Rearrangement. *Cell* **1985**, *42*, 581–588.

- (14) Knowles, P. P.; Murray–Rust, J.; Kjær, S.; Scott, R. P.; Hanrahan, S.; Santoro, M. and Ibáñez, C. F.; McDonald, N. Q. Structure and Chemical Inhibition of the RET Tyrosine Kinase Domain. *J. Biol. Chem.* **2006**, *281*, 33577–33587.
- (15) Yang, F.; He, X.; Feng, L.; Mizuno, K.; Liu, X.; Russell, J.; Xiong, W.; Lu, B. PI–3 kinase and IP3 are both necessary and sufficient to mediate NT3–induced synaptic potentiation. *Nat. Neurosci.* **2001**, *4*, 19–28.
- (16) Santoro, M.; Melillo, R. M.; Carlomagno, F.; Vecchio, G.; Fusco, A. RET : Normal and Abnormal Functions. *Endocrinology* **2004**, *145*, 5448–5451.
- (17) Airaksinen, M. S. and Titievsky, A. and Saarma, M. GDNF family neurotrophic factor signaling: four masters, one servant? *Mol. Cell. Neurosci.* **1999**, *13*, 313–325.
- (18) Tansey, G.; Baloh, R. H.; Milbrandt, J.; Johnson, E. M. GFR $\alpha$  –Mediated Localization of RET to Lipid Rafts Is Required for Effective Downstream Signaling, Differentiation and Neuronal Survival. *Neuron* **2000**, *25*, 611–623.
- (19) Sidorova, Y. A.; Mörtlik, K.; Paveliev, M.; Lindahl, M.; Piranen, E.; Milbrandt, J.; Arumäe, U.; Saarma, M.; Beshpalov, M. M. Persephin signaling through GFR $\alpha$ 1: The potential for the treatment of Parkinson's disease. *Mol. Cell. Neurosci.* **2010**, *44*, 223–232.
- (20) Towbin, H.; Staehelin, T.; Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 4350–4354.
- (21) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3–26.
- (22) The Open Babel Package. version 2.3.1., <http://openbabel.org>.
- (23) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*.
- (24) Fraser, T. R.; Crum–Brown, A. Action of the salts of the ammonium. *Royal Society of Edinburgh.*, 1868.
- (25) Hansch, C.; Fujita, T.  $\rho$ – $\sigma$ – $\pi$  Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *178*, 1616–1626.
- (26) Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inf.* **2010**, *29*, 476–488.
- (27) Karelson, M., *Molecular descriptors in QSAR/QSPR*; Wiley–Interscience: 2000.
- (28) Katritzky, A. R.; Kuanar, M.; Slavov, S.; Hall, C. D.; Karelson, M.; Kahn, I.; Dobchev, D. A. Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. *Chem. Rev.* **2010**, *110*, 5714–5789.

- (29) University of Florida. Gainesville, Florida, USA., *CODESSA PRO User's Manual*, 2005.
- (30) Advanced Chemistry Development, Inc. Toronto, Canada, <http://www.acdlabs.com>.
- (31) Hypercube, Inc. Gainesville, Florida, USA., <http://www.hyper.com>.
- (32) Hocquet, A.; Langg rd, M. An Evaluation of the MM + Force Field. *J. Mol. Modell.* **1998**, *4*, 94–112.
- (33) Stewart, J. J. P. MOPAC: A Semiempirical Molecular Orbital Program. *J. Comput.–Aided Mol. Des.* **1990**, *4*, 1–103.
- (34) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. The Development and Use of Quantum–Mechanical Molecular–Models. 76. AM1 – A New General–Purpose Quantum–Mechanical Molecular–Model. *J. Am. Chem. Soc.*
- (35) Fisher, R. A. On the Mathematical Foundations of Theoretical Statistics. *Philos. Trans. Royal Soc.* **1922**, *222*, 309–368.
- (36) StatSoft, Inc. *Electronic Statistics Textbook*; <http://www.statsoft.com/textbook/>.
- (37) Rohrbach, R. H.; Jurs, P. C. Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. *Anal. Chim. Acta.* **1987**, *199*, 99.
- (38) *Atkins' Physical chemistry*, 9th ed.; Oxford University Press: Oxford, 2010, 602–605.
- (39) Csizmadia, I. G., *Theory and Practice of MO Calculations on Organic Molecules*; Elsevier Scientific Publishing Company: Amsterdam, 1976.
- (40) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum–Chemical Descriptors in QSAR / QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027–1043.
- (41) Tropsha, A.; Gramatica, P.; Gombar, V. K. The Importance of Being Earnest : Validation is the Absolute Essential for Successful Application and Interpretation of QSAR Models. *QSAR Comb. Sci.* **2003**, *22*, 69–77.
- (42) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Ter, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57*, 4977–5010.
- (43) Wold, S; Eriksson, L. Statistical Validation of QSAR Results. *Chemomet. Methods Mol. Des.* **1995**, ed. by van de Waterbeemd, H., 309–318.
- (44) T mm, K.; Fara, D. C.; Katritzky, A. R.; Burk, P.; Karelson, M. A Quantitative Structure–Property Relationship Study of Lithium Cation Basicities. *J. Phys. Chem. A* **2004**, *108*, 4812–4818.
- (45) Kvantitatiivsed Struktuur–Aktiivsus Anal  sid. Mudelite koostamine ja kasutamine EVS 899:2009. Tallinn: Eesti Standardikeskus, 2009.

## Lisad

**Tabel 9.** Kolmeparameetrilise mudeli deskriptorid ja statistilised näitajad.

Nr.	Deskriptori nimi	Koefitsient	Vabaliige	<i>t</i> -test
1	HOMO-1 energy	0,38	0,009	7,75
2	ZX Shadow / ZX Rectangle	−2,43	0,006	−6,47
3	Polarity parameter (Zefirov)	5,68	0,004	4,01
Vabaliige			5,25	9,73

$N = 70$ ;  $n = 3$ ;  $R^2 = 0,70$ ;  $R^2_{cv} = 0,66$ ;  $F = 50,2$ ;  $s^2 = 0,020$

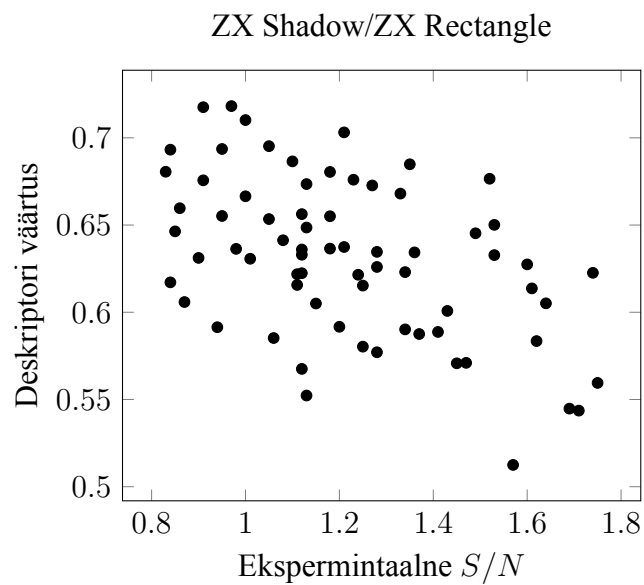
**Tabel 10.** Viieparameetrilise mudeli deskriptorid ja statistilised näitajad.

Nr.	Deskriptori nimi	Koefitsient	Vabaliige	<i>t</i> -test
1	ZX Shadow / ZX Rectangle	−2,68	0,05	−8,60
2	Max partial charge (Zefirov) for all atom types	6,23	0,33	8,57
3	Internal heat capacity (300K) /natoms	−1,59	0,52	−6,51
4	HOMO-1 energy	0,26	0,28	5,61
5	HA dependent HDSA-2 (MOPAC PC)	0,008	0,10	3,68
Vabaliige			7,46	16,50

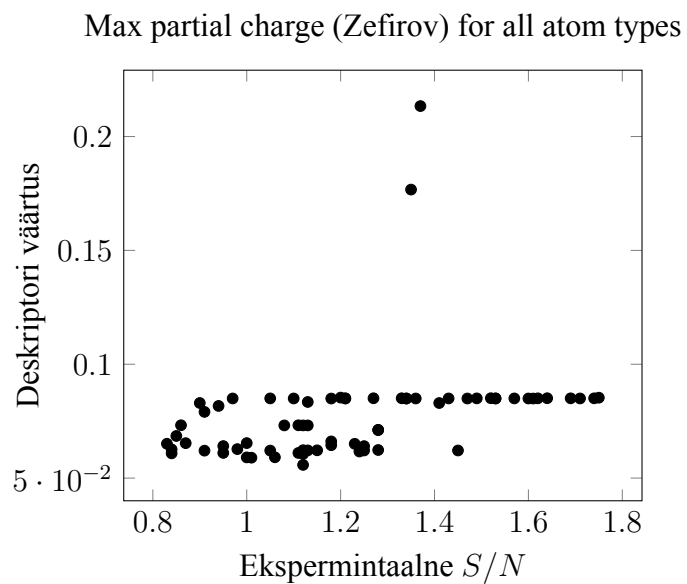
$N = 70$ ;  $n = 5$ ;  $R^2 = 0,81$ ;  $R^2_{cv} = 0,77$ ;  $F = 52,9$ ;  $s^2 = 0,013$

$N$  – ühendite arv,  $n$  - deskriptorite arv,  $R^2$  - korrelatsioonikoefitsient,  $R^2_{cv}$  - ristvalideeritud korrelatsioonikoefitsient,  $F$  - Fisheri kriteerium,  $s^2$  - standardhälbe ruut.

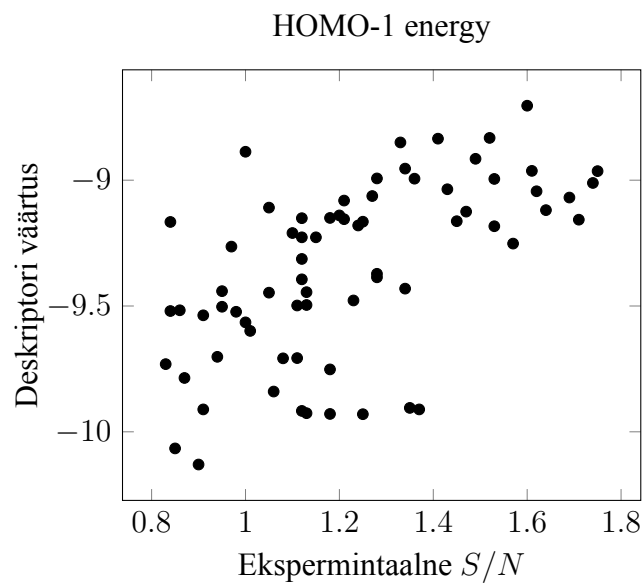




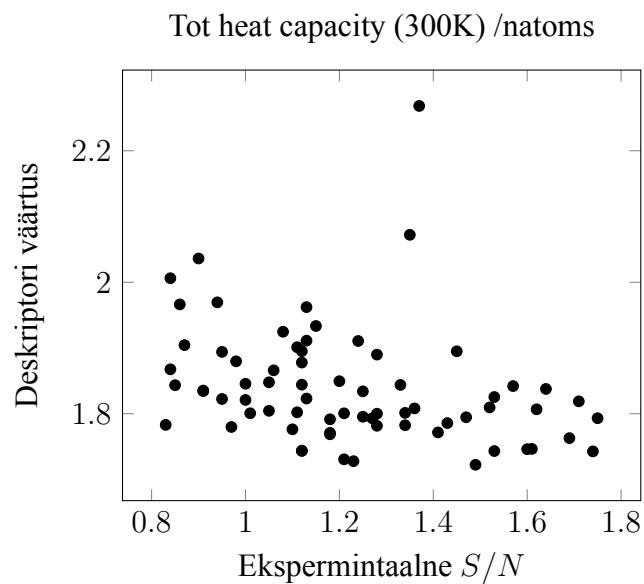
**Joonis 5.** ZX varju/ZX ristküliku ja eksperimentaalsete  $S/N$  väärtuste suhe.



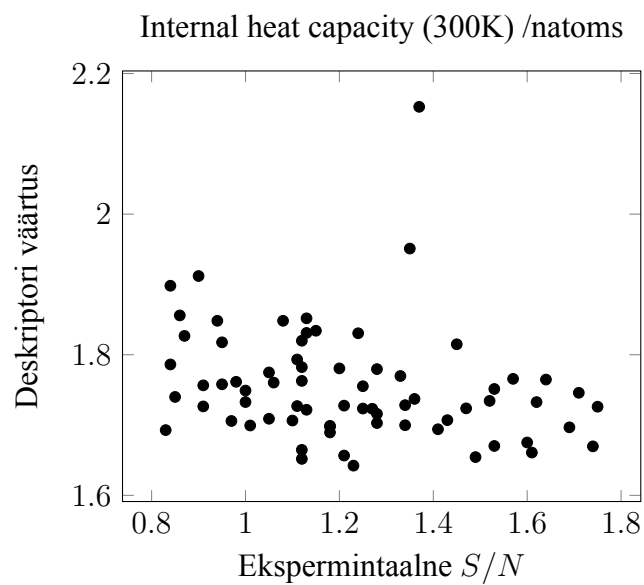
**Joonis 6.** Maksimaalse osalaengu (Zefirov) kõigi aatomitüüpide jaoks ja eksperimentaalsete  $S/N$  väärtuste suhe.



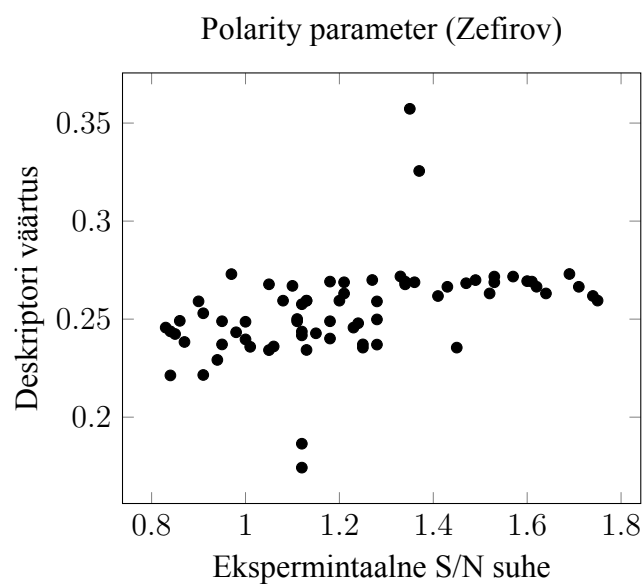
**Joonis 7.** HOMO-1 energia ja eksperimentaalsete  $S/N$  väärtuste suhe.



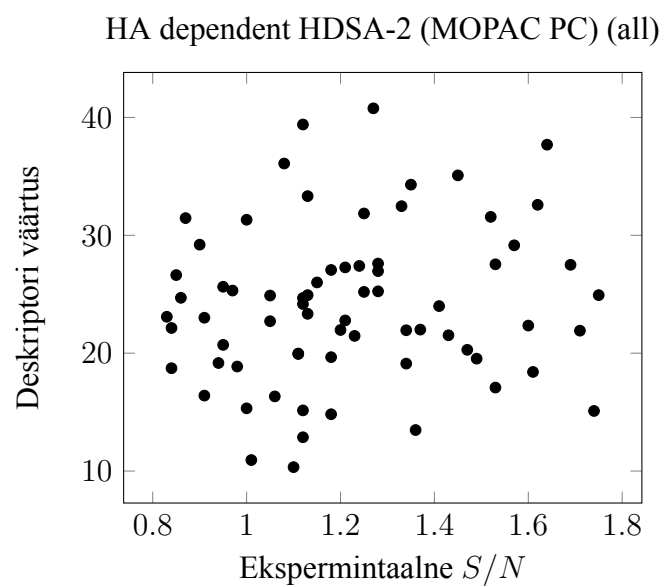
**Joonis 8.** Täieliku soojusmahtuvuse (300K)/ aatomite arv ja eksperimentaalsete  $S/N$  väärtuste suhe.



**Joonis 9.** Sisemise soojusmahtuvuse (300K)/ aatomite arv ja eksperimentaalsete  $S/N$  väärtuste suhe.

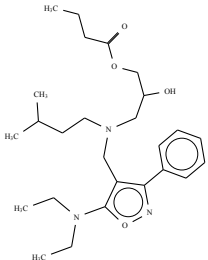
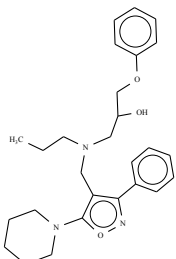
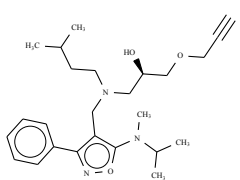
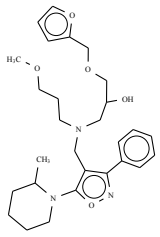
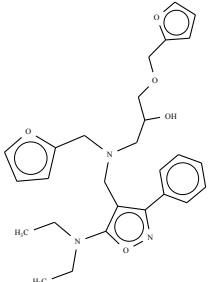


**Joonis 10.** Polaarsuse parameeter (Zefirov) ja eksperimentaalsete  $S/N$  väärtuste suhe.



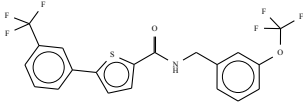
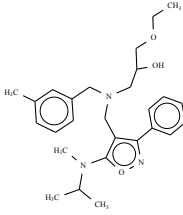
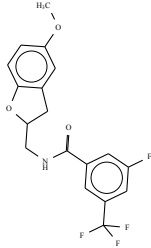
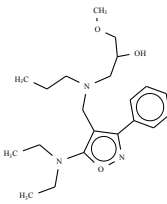
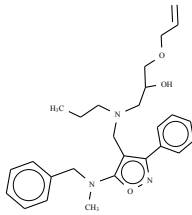
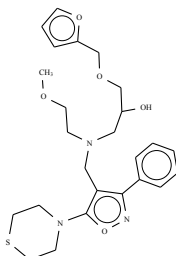
**Joonis 11.** Vesiniksideme doonori pindala ja eksperimentaalsete  $S/N$  väärtuste suhe.

**Tabel 11.** Ühendite eksperimentaalsed ja ennustatud väärtused.

Ühendi number	Struktuur	Eksperimentaalne	Ennustatud	Erinevus
1		1,75	1,61	0,14
2		1,74	1,50	0,24
3		1,71	1,57	0,14
4		1,69	1,66	0,03
5		1,64	1,40	0,24

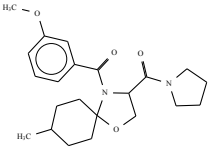
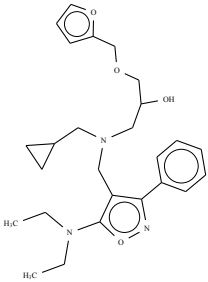
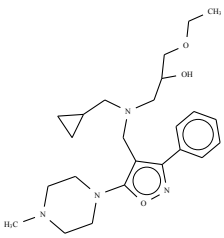
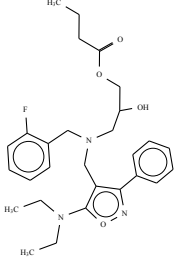
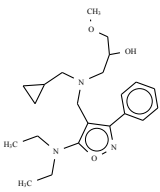
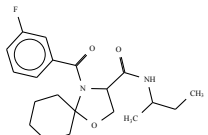
Ühendi number	Struktuur	Eksperimentaalne	Ennustatud	Erinevus
6		1,62	1,51	0,11
7		1,61	1,53	0,08
8		1,6	1,56	0,04
9		1,57	1,60	-0,03
10		1,53	1,28	0,25
11		1,53	1,47	0,06

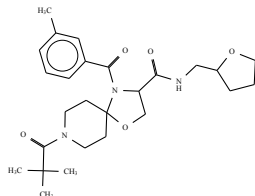
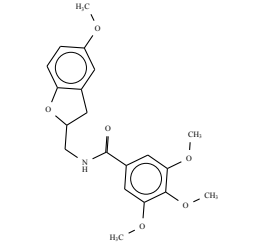
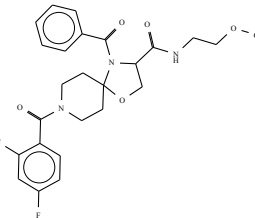
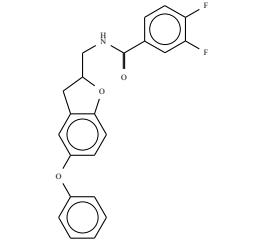
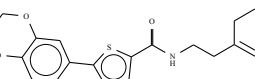
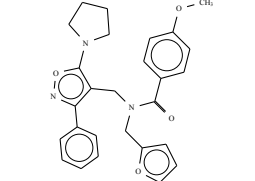


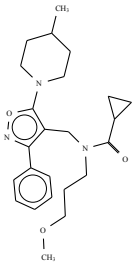
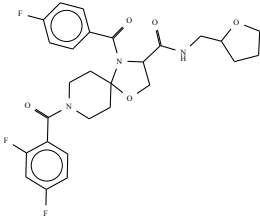
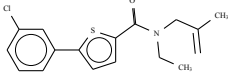
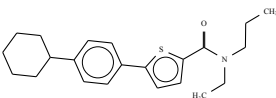
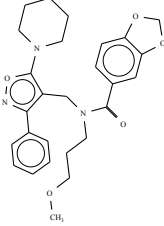
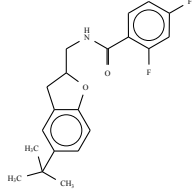
Ühendi number	Struktuur	Eksperimentaalne	Ennustatud	Erinevus
18		1,37	1,47	-0,10
19		1,36	1,39	-0,03
20		1,35	1,24	0,11
21		1,34	1,34	0,00
22		1,34	1,52	-0,18
23		1,33	1,29	0,04

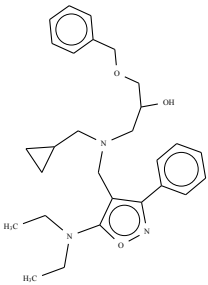
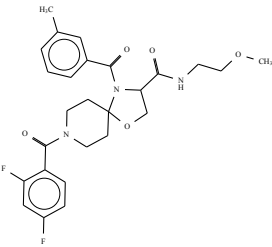
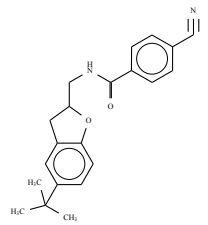
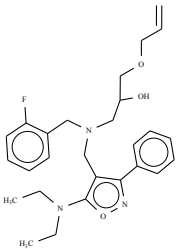
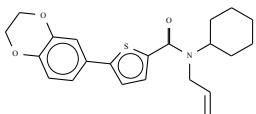
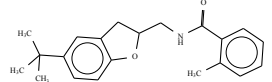


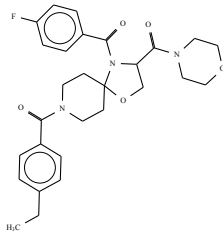
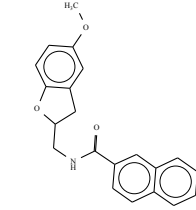
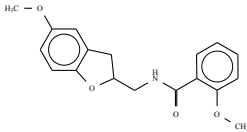
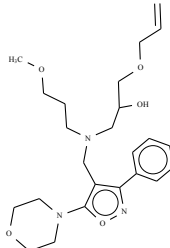
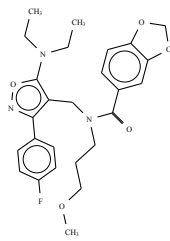
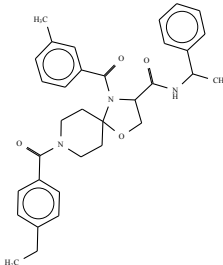
Ühendi number	Struktuur	Eksperimentaalne	Ennustatud	Erinevus
24		1,28	1,39	-0,11
25		1,28	1,31	-0,03
26		1,28	1,08	0,20
27		1,27	1,29	-0,02
28		1,25	1,22	0,03
29		1,25	1,19	0,06
30		1,24	1,10	0,14

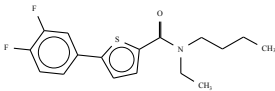
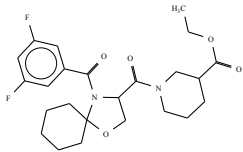
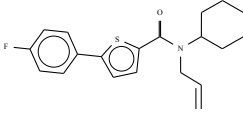
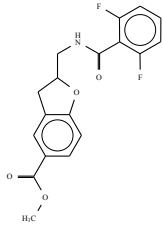
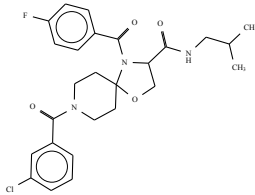
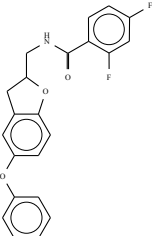
Ühendi number	Struktuur	Eksperimentaalne	Ennustatud	Erinevus
31		1,23	1,14	0,09
32		1,21	1,37	-0,16
33		1,21	1,26	-0,05
34		1,2	1,41	-0,21
35		1,18	1,27	-0,09
36		1,18	1,06	0,12

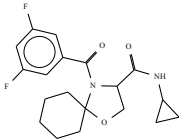
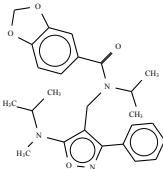
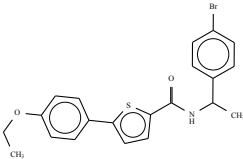
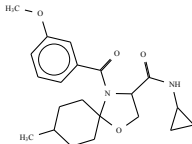
Ühendi number	Struktuur	Eksperimentaalne	Ennustatud	Erinevus
37		1,18	1,07	0,11
38		1,15	1,11	0,04
39		1,13	1,17	-0,04
40		1,13	0,95	0,18
41		1,13	1,08	0,05
42		1,12	1,13	-0,01

Ühendi number	Struktuur	Eksperimentaalne	Ennustatud	Erinevus
43		1,12	1,20	-0,08
44		1,12	1,01	0,11
45		1,12	1,30	-0,18
46		1,12	1,19	-0,07
47		1,11	1,16	-0,05
48		1,11	1,07	0,04

Ühendi number	Struktuur	Eksperimentaalne	Ennustatud	Erinevus
49		1,1	1,23	-0,13
50		1,08	0,97	0,11
51		1,06	1,08	-0,02
52		1,05	1,26	-0,21
53		1,05	0,98	0,07
54		1,01	1,10	-0,09

Ühendi number	Struktuur	Eksperimentaalne	Ennustatud	Erinevus
55		1	0,91	0,09
56		1	1,13	-0,13
57		0,98	1,03	-0,05
58		0,97	1,13	-0,16
59		0,95	0,97	-0,02
60		0,95	0,96	-0,01

Ühendi number	Struktuur	Eksperimentaalne	Ennustatud	Erinevus
61		0,94	1,10	-0,16
62		0,91	0,98	-0,07
63		0,91	0,86	0,05
64		0,9	0,81	0,09
65		0,87	1,02	-0,15
66		0,86	0,92	-0,06

Ühendi number	Struktuur	Eksperimentaalne	Ennustatud	Erinevus
67		0,85	0,94	-0,09
68		0,84	0,97	-0,13
69		0,84	0,92	-0,08
70		0,83	0,99	-0,16



## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Karl Normak,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose  
“GDNF mimeetikute GFR $\alpha$ 1-RET aktiveerimise eksperimentaalne uurimine ja QSAR  
modelleerimine”,

mille juhendajad on Lauri Sikk, Kaido Tamm,

- 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
- 1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu alates **02.06.2019** kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **23.05.2017**